



REC'D 06 AUG 2004
WIPO
PCT

DK/04/478

Kongeriget Danmark

BEST AVAILABLE COPY

Patent application No.: PA 2003 01010

Date of filing: 03 July 2003

Applicant:
(Name and address)
Aarhus Universitet
Nordre Ringgade 1
DK-8000 Århus C
Denmark

Title: Nod-factor perception

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.



Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

28 July 2004

Susanne Morsing
Susanne Morsing

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

03 JULI 2003

1

Modtaget

Nod-factor perception

Field of the invention

The invention relates to a novel Nod-factor binding element and component
5 polypeptides that are useful in enhancing Nod-factor binding in nodulating
plants and inducing nodulation in non-nodulating plants. More specifically,
the invention relates to Nod-factor binding proteins and their respective
genomic and mRNA nucleic acid sequences.

10 Background of the invention

The growth of agricultural crops is almost always limited by the availability of
nitrogen, and at least 50% of global needs are met by the application of
synthetic fertilisers in the form of ammonia, nitrate or urea. Apart from
recycling of crop residues and animal manure, and atmospheric deposition,
15 the other most important source of nitrogen for agriculture comes from
biological nitrogen fixation.

A small percentage of prokaryotes, the diazotrophs, produce nitrogenases and
are capable of nitrogen fixation. Members of this group, belonging to the
20 *Rhizobiaceae* family (for example *Mesorhizobium loti*, *Rhizobium meliloti*,
Bradyrhizobium japonicum, *Rhizobium leguminosarum* bv *viceae*) here
collectively called *Rhizobium* or *Rhizobia* spp and the actinobacterium
Frankia spp, can form endosymbiotic associations with plants conferring the
ability to fix nitrogen. Although many plants can associate with nitrogen fixing
bacteria, only a few plants, all members of the Rosid I Clade, form
25 endosymbiotic associations with *Rhizobia* spp and *Frankia* spp., which are
unique in that most of the nitrogen is transferred to and assimilated by the
host plant. Legumes, including soybean, bean, pea, peanut, chickpea,
cowpea, lentil, pigeonpea, alfalfa and clover, are the most agronomically
important members of this small group of nitrogen-fixing plants.
30 The rhizobial-legume interaction is generally host-strain specific, whereby
successful symbiotic associations only occur between specific rhizobial

strains and a limited number of legume species. The specificity of this interaction is determined by chemical signalling between plant and bacteria, which accompanies the initial interaction and the establishment of the symbiotic association (Hirsch *et al.* 2001, *Plant Physiol.* 127: 1484-1492).

5 Specific (iso)flavanoids, secreted into the soil by legume spp, allow *Rhizobium* spp to distinguish compatible hosts in their proximity and to migrate and associate with roots of the host. In a compatible interaction, the (iso)flavanoid perceived by the *Rhizobium* spp, interacts with the rhizobial *nodD* gene product, which in turn leads to the induction of rhizobial Nod-factor synthesis. Nod-factor molecules are lipo-chitin-oligosaccharides, commonly comprising four or five β -1-4 linked N-acetylglucosamines, with a 16 to 18 carbon chain fatty acid n-acetylated on the terminal non-reducing sugar. Nod factors are synthesised in a number of variants, characterised by their chemically different substitutions on the chitin backbone which are

10 distinguished by the compatible host plant. The perception of Nod-factors by the host induces invasion zone root hairs, in the proximity of rhizobial cells, to curl and entrap the bacteria. The adjacent region of the root hair plasma membrane invaginates and new cell wall material is synthesized to form an infection thread or tube, which serves to transport the symbiotic bacteria

15 through the epidermis to the cortical cells of the root. Here the cortical cells are induced to divide to form a primordium, from which a root nodule subsequently develops. In legumes belonging to genera like *Arachis* (peanut), *Stylosantos* and *Sesbania*, infection is initiated by a simple "crack entry" through spaces or cavities between epidermal cells and lateral roots.

20 25 In spite of these differences, perception of Nod factors by the host plant simultaneously induces the expression of a series of plant nodulin genes, which control the development and function of root nodules, wherein the rhizobial endosymbiotic association and nitrogen fixation are localised. A variety of molecular approaches have identified a series of plant nodulin

30 genes which play a role in rhizobial-legume symbiosis, and whose expression is induced at early or later stages of rhizobial infection and nodule

development (Geurts and Bisseling, 2002, *Plant Cell* supplement S239-249). Furthermore, plant mutant studies have revealed that a signalling pathway must be involved in amplifying and transducing the signal resulting from nod-factor perception, which is required for the induction of nodulin gene

5 expression. Among the first physiological events identified in this signal transduction pathway, which occurs circa 1 min after Nod-factor application to the root epidermis, is a rapid calcium influx followed by chloride efflux, causing depolarisation of the plasma membrane and alkalinization of the external root hair space of the invasion zone. A subsequent efflux of

10 potassium ions allows re-polarisation of the membrane, and later a series of calcium oscillations are seen to propagate the signal through the root hair cell. Pharmacological studies with specific drugs, which mimic or block Nod-factor induced responses, have identified potential components of the signalling pathway. Thus mastoparan, a peptide which is thought to mimic

15 the activated intracellular domain of G-protein coupled receptors, can induce early Nod gene expression and root hair curling. This suggests that trimeric G protein may be involved in the Nod-factor signal transduction pathway. Analysis of a group of nodulation mutants, including some that fail to show calcium oscillations in response to Nod-factor signals, has revealed that in

20 addition to the lack of nodulation, these mutants are unable to form endosymbioses with arbuscular mycorrhizal fungi. This implies that a common symbiotic signal transduction pathway is shared by two types of endosymbiotic relationships, namely root nodule symbiosis, which is largely restricted to the legume family, and arbuscular mycorrhizal symbiosis, which

25 is common to the majority of land plant species. This suggests that there may be a few key genes which dispose legumes to engage in nodulation, and which are missing from crop plants such as cereals.

The identification of these key genes, which encode functions which are indispensable for establishing a nitrogen fixing system in legumes, and their

30 transfer and expression in non-nodulating plants, has long been a goal of molecular plant breeders. This could have a significant agronomic impact on

the cultivation of cereals such as rice, where production of two harvests a year may require fertilisation with up to 400 kg nitrogen per hectare. In accordance with this goal, WO02102841 describes the gene encoding the NORK polypeptide, isolated from the nodulating legume *Medicago sativa*,

5 and the transformation of this gene into plants incapable of nitrogen fixation. The NORK polypeptide and its homologue/orthologue SYMRK from *Lotus japonicus* (Stracke et al 2002 *Nature* 417:959-962), are transmembrane receptor-like kinases with an extracellular domain comprising leucine-rich repeats, and an intracellular protein kinase domain. *Lotus japonicus* mutants,

10 with a non-functional SYMRK gene, fail to form symbiotic relationships with either nodulating rhizobia or arbuscular mycorrhiza. This implies that a common symbiotic signalling pathway mediates these two symbiotic relationships, where SYMRK comprises an early step in the pathway. The *symRK* mutants retain an initial response to rhizobial infection, whereby the

15 root hairs in the susceptible invasion zone undergo swelling of the root hair tip and branching, but fail to curl. This suggests that the SYMRK protein is required for an early step in the common symbiotic signalling pathway, located downstream of the perception and binding of microbial signal molecules (e.g. Nod-factors), that leads to the activation of nodulin gene expression.

20 The search for key symbiosis genes has also focussed on 'candidate genes' encoding receptor proteins with the potential for perceiving and binding Nod-factors or surface structures on rhizobial bacteria. US 6,465,716 discloses NBP46, a Nod-factor binding lectin isolated from *Dolichos biflorus* roots, and

25 its transgenic expression in transformed plants. Transgenic expression of NBP46 in plants is reported to confer the ability to bind to specific carbohydrates in the rhizobial cell wall and thereby to bind these bacteria and utilise atmospheric nitrogen, as well as conferring apyrase activity. An alternative approach to search for key symbiosis genes has been to screen

30 for Nod-factor binding proteins in protein extracts of plant roots. NFBS1 and NFBS2 were isolated from *Medicago trunculata* and shown to bind Nod-

factors in nanomolar concentrations, however, they both failed to exhibit the Nod-factor specificity characteristic of rhizobial-legume interactions (Geurts and Bisseling, 2002 *supra*).

5 The Nod-factor binding element, which is responsible for strain specific Nod-factor perception is not, as yet, identified. The isolation and characterisation of this element and its respective gene(s) would open the way to introducing Nod-factor recognition into non-nodulating plants and thereby the potential to establish *Rhizobium*-based nitrogen fixation in important crop plants.

10

Rhizobial strains produce strain-specific Nod-factors, lipochitin oligosaccharides (LCOs), which are required for a host-specific interaction with their respective legume hosts. *Lotus* and peas belong to two different cross-inoculation groups, where *Lotus* develops nodules after infection with *Mesorhizobium loti*, while pea develops nodules with *Rhizobium leguminosarum* bv *viciae*. Cultivars belonging to a given *Lotus* sp also vary in their ability to interact and form nodules with a given rhizobial strain.

15 Perception of Nod-factor secreted by *Rhizobium* spp bacteria, as the first step in nodulation, commonly leads to the initiation of tens or even hundreds

20 of rhizobial infection sites in a root. However, the majority of these infections abort and only in a few cases do the rhizobia infect the nodule primordium.

25 The frequency and efficiency of the *Rhizobium*-legume interaction leading to infection is known to be influenced by variations in Nod-factor structure. The genetics of Nod-factor synthesis and modification of their chemical structure

in *Rhizobium* spp have been extensively characterised. An understanding of Nod-factor binding and perception, and the structure of its component elements is needed in order to optimise the host Nod-factor response. This information would, in turn, provide the necessary tools to breed for enhanced efficiency of nodulation and nitrogen fixation in current nitrogen-fixing crops.

30

The importance of this goal is clearly illustrated by the performance of the major US legume crop, soybean, which is grown on 15%, or more, of agricultural land in the US. While nitrogen fixation by soybean root nodules can assimilate as much as 100 kg nitrogen per hectare per year, these high

5 levels of nitrogen assimilation are insufficient to support the growth of the highest yielding modern soybean cultivars, which still require the application of fertiliser.

In summary, there is a need to increase the efficiency of nodulation and

10 nitrogen fixation in current legume crops as well as to transfer this ability to non-nodulating crops in order to meet the nutritional needs of a growing global population, while minimising the future use of nitrogen fertilisers and their associated negative environmental impact.

Summary of the invention

15 The invention provides a Nod-factor binding element comprising one or more isolated NFR polypeptides. The NFR polypeptides of the invention are NFR1, comprising an amino acid sequence substantially identical to SEQ ID No: 25, having specific Nod-factor binding properties, and NFR5 comprising an amino acid sequence substantially identical to SEQ ID No: 8, having specific

20 Nod-factor binding properties. Furthermore, the invention provides for the isolation of nucleic acid molecules comprising *NFR1* and *NFR5* gene and cDNA sequences encoding said NFR1 polypeptide and said NFR5 polypeptide, comprising a nucleic acid sequence substantially identical to SEQ ID No: 23 and SEQ ID No: 7, respectively.

25 According to a further embodiment of the invention, a method is provided for producing a plant expressing the Nod-factor binding element, the method comprising introducing into the plant a transgenic expression cassette comprising a nucleic acid sequence, encoding a NFR polypeptide having specific Nod-factor binding properties and having an amino acid sequence

30 substantially identical to SEQ ID No: 25 or SEQ ID No: 8, respectively, wherein the nucleic acid sequence is operably linked to its own promoter or a

heterologous promoter, preferably a root specific promoter. In a preferred embodiment, the expression of both said NFR 1 and NFR5 polypeptides by the transgenic plant confers on the plant the ability to bind Nod-factors in a chemically specific manner and thereby initiate the establishment of a

5 *Rhizobium*-plant interaction leading to the development of nitrogen-fixing root nodules.

According to a further embodiment, the invention provides a method for marker assisted breeding of *NFR* alleles, encoding variant *NFR* polypeptides, comprising the steps of identifying variant *NFR1* or *NFR5* polypeptides in a

10 nodulating legume species, comprising an amino acid sequence substantially similar to SEQ ID No: 25 or SEQ ID No: 8 respectively; determining the nodulation frequency of plants expressing said variant *NFR1* or *NFR5* polypeptide; identifying DNA polymorphisms at loci genetically linked to or within the allele locus encoding said variant *NFR1* or *NFR5* locus; preparing

15 molecular markers based on said DNA polymorphisms; and using said molecular markers for the identification and selection of plants carrying *NFR* alleles encoding said variant *NFR1* or *NFR5* polypeptides. The invention includes plants selected by the use of this method of marker assisted breeding. In a preferred embodiment, said method of marker assisted

20 breeding of *NFR* alleles provides for the breeding legumes with enhanced nodulation frequency and nodule occupancy.

Brief Description of the figures

Figure 1: Map based cloning of *Lotus NFR5*. a. Genetic map of the *NFR5* region with positions of linked AFLP and microsatellite markers above the line and distances in cM below. The fraction of recombinant plants detected in the mapping population is indicated. b. Physical map of the BAC and TAC clones between the closest linked microsatellite markers. The positions of sequence-derived markers used to fine-map the *NFR5* locus, and the fraction of recombinant plants found in the mapping population are indicated. c. Candidate genes identified in the sequenced region delimited by the closest

linked recombination events. **d.** Structure of the *NFR5* gene, position of the transcription initiation point and the *nfr5-1*, *nfr5-2* and *nfr5-3* mutations. The asterisk indicates the position of a stop codon in *nfr5-3*; the black triangle a retrotransposon insertion in *nfr5-2*; and the grey box defines the deletion in **5** *nfr5-1*. GGDP: geranylgeranyl diphosphate synthase; RE: retroelement; RZF: ring zinc finger protein; GT: glycosyl transferase; A2L: apetala2-like protein; RLK: receptor-like kinase; PL: pectate lyase-like protein; AS: ATPase-subunit; HD: homeodomain protein; RF: ring finger protein. Hypothetical proteins are not labelled. **e.** Southern hybridization demonstrating deletion of **10** *SYM10* in the "N15" *sym10* mutant line. *EcoRI* digested genomic DNA of the parental variety "Sparkle" and the fast neutron derived mutant "N15" hybridized with a pea *SYM10* probe covering the region encoding the predicted extracellular domain. Hybridization with a probe from the 3' untranslated region demonstrated that the complete gene was deleted.

15 **Figure 2:** Structure and domains of the *NFR5* protein. **a.** Schematic representation of the *NFR5* protein domains. **b.** The amino acid sequence of *NFR5* arranged in protein domains. Bold, conserved LysM residues. Bold and underlined residues conserved in protein kinase domains (KD); TM: transmembrane, SP: signal peptide. The asterisk indicates a stop codon in **20** the *nfr5-3*; the black triangle a retrotransposon insertion in *nfr5-2* and the grey box defines the amino acids deleted in *nfr5-1*. **c.** Individual alignment of the three LysM motifs (M1, M2, M3) from *NFR5*, pea *SYM10*, *Medicago truncatula* (*M.t*, Ac126779) rice (Ac103891), the single LysM in chitinase from *Volvox carteri* (Acc. No: T08150) and the pfam consensus. **d.** The divergent or absent activation loop (domain VIII) in the *NFR5* family of receptor kinases is illustrated by alignment of kinase motifs VII, VIII and IX from *Arabidopsis* (At2g33580) *NFR5*, *SYM10*, *Medicago truncatula* (*M.t*, Ac126779), rice (Ac103891) and the SMART consensus. Conserved domain VII aspartic acid is marked in bold and underlined. **c** and **d** the amino acids conserved in all **25**

aligned motifs are marked in black and amino acids conserved in two or more motifs are marked in grey.

Figure 3. The aligned amino acid sequence of the *LjNFR5* and *PsSYM10* proteins. Amino acid residues sharing identity are highlighted. The *Medicago truncatula* (Ac126779) showing 76 % amino acid identity to *Lotus* NFR5 is included to exemplify a substantially identical protein sequence.

Figure 4. Steady-state levels of *LjNFR5* and *PsSYM10* mRNA. a. *NFR5* mRNA detected in uninoculated roots, inoculated roots, nodules, leaves, flowers and pods of *Lotus* plants. b. Time course of *NFR5* mRNA transcript accumulation in roots after inoculation with *M. loti*. The identity of the amplified transcripts was confirmed by sequencing. ATPase was used as internal control and relative normalised values compared to uninoculated roots are shown. c. Northern analysis showing *NFR5* mRNA expression in nodule leaf and root of symbiotically and non-symbiotically grown *Lotus* plants. d. Northern analysis showing *Sym10* mRNA expression in leaf, root and nodule of symbiotically and non-symbiotically grown pea plants.

Figure 5. Positional cloning of the *NFR1* gene. a. Genetic map of the region surrounding the *NFR1* locus. Positions of the closest AFLP, microsatellite- and PCR-markers are given together with genetic distances in cM. b. Physical map of the *NFR1* locus. BAC clones 56L2, 16K18, 10M24, 36D15, 56K22 and TAC clones LjT05B16, LjT02D13, LjT211O02, which cover the region are shown. The numbers of recombination events detected with BAC and TAC end-markers or internal markers are given. Arrows indicate the positions of the two markers (10M24-2, 56L2-2) delimiting the *NFR1* locus. c. UFD and HP correspond to the UFD1-like protein and the hypothetical protein encoded in the region. c. Exon-intron structure of the *NFR1* gene. Boxes correspond to exons, where LysM motifs are shown in light grey, trans-membrane region in black, kinase domains in dark grey. Dotted lines define introns and full lines define the 5' and 3' un-translated regions. The

nucleotide length of all exons and introns are indicated. The numbers between brackets correspond to exon and intron 4, corresponding to alternative splicing.

Figure 6. Structure and domains of the NFR1 protein. a. Primary structure of the *NFR1* protein comprising a signal peptide (SP); LysM motifs (LysM1 and LysM2); transmembrane region (TM); protein kinase domains with conserved amino acids in bold and underlined (PK). The cysteine couples (CxC) are in bold and the LysM amino acids important for secondary structure maintenance are underlined. The two extra amino acids resulting from alternative splicing are shown in brackets. I-XI represent the kinase domains. Asterisks indicate positions of the nonsense mutations found in *NFR1-1* and *NFR1-2* mutant alleles. **b.** Alignments of the two *NFR1* LysM motifs to the consensus sequences predicted by the SMART program and the *Arabidopsis thaliana* (Acc No: NP566689), rice (*O. sativa*) (Acc No: BAB89226), and *Volvox carteri* (Acc. No: T08150) LysM motifs.

Figure 7. *NFR1*, *NFR5* and *SymRK* gene expression. a. Transcript levels of *NFR1* in uninoculated, inoculated roots, nodules, leaves, flowers and pods of wild type plants. **b.** Transcript levels of *NFR1* in wild type, *nfr1*, *nfr5* and *symRK* mutant plants after inoculation with *M. loti*. **c.** Transcript levels of *NFR5* in wild type, *nfr1*, *nfr5* and *SymRK* mutant plants after inoculation with *M. loti*. **d.** Transcript levels of *SYMRK* in wild type, *nfr1*, *nfr5* and *symRK* mutant plants after inoculation with *M. loti*. Transcript levels were measured by quantitative PCR. ATPase was used as internal control and relative values normalised to the untreated roots (zero hours) are shown.

Figure 8. Root hair response after inoculation with *M. loti* or Nod-factor application. a. Wild type root hair curling on seedlings inoculated with *M. loti*. **b.** Root hair deformations on wild type seedlings after Nod-factor application. **c.** Root hairs on *nfr1-1* seedlings inoculated with *M. loti*. **d.** Root hairs on

nfr1-1 seedlings after Nod-factor application. e. Root hairs with balloon deformations on *symRK-3* mutants inoculated with *M. loti*. f. Roots hairs on a *nfr1-1,symRK-3* double mutant inoculated with *M. loti* g. Excessive root hair response on *nin* mutants inoculated with *M. loti*. h. Root hairs on a *nfr1-1,nin* double mutant inoculated with *M. loti*. Root hairs on *nfr5-1* seedlings inoculated with *M. loti*, *nfr5-1* seedlings after Nod-factor application, untreated *nfr5-1* control, untreated wild type control, untreated *nfr1-1* control, are indistinguishable from the straight roots hairs shown in c, d, f, h and therefore not shown. Inserts to the right of a to h show a close-up of the root hairs.

Figure 9. Membrane depolarisation and pH changes in the extracellular root hair space after application of Nod-factor purified from *M. loti*. Influence of 0.1 μ M Nod-factor (NF) on membrane potential (Em) and/or external pH (pH) of a. *Lotus* wild type b. *nfr5-1* and *nfr5-2* mutants c. *nfr1-1* and *nfr1-2* mutants d. *symRK-1* and *symRK-3* mutants e. *nfr1-2,symRK-3* double mutant, f. pH changes in the extracellular root hair space after application of an undecorated chito-octaose.

Figure 10. Expression of the *NIN* and *ENOD2* genes in wild type, *nfr1* and *nfr5* mutant genotypes. a. *NIN* transcript level in RNA extracted from roots two hours to 12 days after *M. loti* inoculation. b *ENOD2* transcript level in RNA extracted from roots two hours to 12 days after *M. loti* inoculation. Transcript levels were measured by quantitative PCR and the identity of the amplified sequences was confirmed by sequencing. ATPase was used as internal control and relative values normalised to the untreated root (zero hours) are shown.

Figure 11. Alignment NFR1 and NFR5 proteins reveal an overall similarity of 33 % amino acid identities

Figure 12. Domain structure of native and hybrid NFR1 and NFR5 polypeptides.

Detailed description of the invention

5

I. Definitions

AFLP: Amplified Fragment Length Polymorphism is a PCR-based technique for the amplification of genomic fragments obtained after digestion with two different enzymes. Different genotypes can be differentiated based on the 10 size of amplified fragments or by the presence or absence of a specific fragment (Vos, P. (1998), *Methods Mol Biol.*, 82:147-155). Amplified Fragment Length Polymorphism is a PCR-based technique used to map genetic loci.

Agrobacterium rhizogenes-mediated transformation: is a technique used 15 to obtain transformed roots by infection with *Agrobacterium rhizogenes*. During the transformation process the bacteria transfers a DNA fragment (T-DNA) from an endogenous plasmid into the plant genome (Stougaard, J. et al, (1987) *Mol. Gen. Genet.* 207, 251-255). For transfer of a gene of interest the gene is first inserted into the T-DNA region of *Agrobacterium rhizogenes* 20 which is subsequently used for wound-site infection.

Allele: gene variant

BAC clones: clones from a Bacterial Artificial Chromosome library

Conservatively modified variant: when referring to a polypeptide sequence when compared to a second sequence, includes individual conservative 25 amino acid substitutions as well as individual deletions, or additions of amino acids. Conservative amino acid substitution tables, providing functionally similar amino acids are well known in the art. When referring to nucleic acid sequences, conservative modified variants are those that encode an identical amino acid sequence, in recognition of the fact that codon redundancy allows 30 a large number of different sequences to encode any given protein.

Contig: a series of overlapping cloned sequences e.g. BACs, co-linear and homologous to a region of genomic DNA.

Exons: protein coding sequences of a gene sequence

Expression cassette: refers to a nucleic acid sequence, comprising a

5 promoter operably linked to a second nucleic acid sequence containing an ORF or gene, which in turn is operably linked to a terminator sequence.

Heterologous: A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or from a different gene, or is modified from its original form. A heterologous

10 promoter operably linked to a coding sequence refers to a promoter from a species, different from that from which the coding sequence was derived, or, from a gene, different from that from which the coding sequence was derived.

Homologue: is a gene or protein with substantial identity to another gene's sequence or another protein's sequence.

15 **Identity:** refers to two nucleic acid or polypeptide sequences that are the same or have a specified percentage of nucleic acids or amino acids that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the sequence comparison algorithms listed herein, or by manual alignment and visual inspection. When

20 percentage of sequence identity is used in reference to proteins, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional

25 properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to account for the conservative nature of the substitution. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thus increasing the percent identity. Means for making these adjustments are

30 well known to those skilled in the art.

Introns: are non-coding sequences interrupting protein coding sequences within a gene sequence.

LCO: lipochitin oligosaccharides.

Legumes: are members of the plant Family Fabaceae, and include bean, pea, soybean, clover, vetch, alfalfa, peanut, pigeon pea, chickpea, fababean, cowpea, lentil in total approximately 20.000 species.

Locus: or "loci" refers to the map position of a nucleic acid sequence or gene on a genome.

Marker assisted breeding: the use of DNA polymorphisms as "molecular markers", (for examples simple sequence repeats (microsatellites) or single nucleotide polymorphism (SNP)) which are found at loci, genetically linked to, or within, the *NFR1* or *NFR5* loci, to breed for advantageous *NFR* alleles.

Molecular markers: refer to sites of variation at the DNA sequence level in a genome, which commonly do not show themselves in the phenotype, and

5 **may be a single nucleotide difference in a gene, or a piece of repetitive DNA.**

Monocotyledenous cereal: includes, but is not limited to, barley, maize, oats, rice, rye, sorghum, and wheat.

Mutant: a plant or organism with a modified genome sequence resulting in a phenotype which differs from the common wild-type phenotype.

10 **Native:** as in "native promoter" refers to a promoter operably linked to its homologous coding sequence.

NFR : refers to *NFR* genes, in particular *NFR1* and *NFR5* genes which encode *NFR1* and *NFR5* polypeptides respectively, and comprise a nucleic acid sequence substantially identical to SEQ ID No: 23 and SEQ ID No: 7,

15 respectively.

NFR polypeptides: are polypeptides that are required for Nod-factor binding and function as the Nod-factor binding element in nodulating plants. *NFR* polypeptides include the *NFR5* polypeptide, having an amino acid sequence substantially similar to SEQ ID No: 8, and the *NFR1* polypeptide, having an

20 amino acid sequence substantially similar to SEQ ID No: 25. *NFR5* and *NFR1* polypeptides show little sequence homology, but they share a similar

domain structure comprising an N-terminal signal peptide, an extracellular domain having 2 or 3 LysM-type motifs, followed by a transmembrane domain, followed by an intracellular domain comprising a kinase domain characteristic of serine/threonine kinases. The extracellular domain of NFR

5 proteins is the primary determinant of the specificity of Nod-factor recognition, whereby a host plant comprising a given *NFG* allele will only form nodules with one or a limited number of *Rhizobium* strains.

Northern blot analysis: a technique for the quantitative analysis of mRNA species in an RNA preparation.

10 **Nod-factors:** are synthesised by nitrogen-fixing *Rhizobium* bacteria, which form symbiotic relationships with specific host plants. They are lipo-chitin-oligosaccharides (LCOs), commonly comprising four or five β -1-4 linked N-acetylglucosamines, with a 16 to 18 carbon chain fatty acid n-acetylated on the terminal non-reducing sugar. Nod-factors are synthesised in a number of

15 chemically modified forms, which are distinguished by the compatible host plant.

Nod-factor binding element: comprises one or more NFR polypeptides present in the roots of nodulating plants, and functions in detecting the presence of Nod-factors at the root surface and within the root and nodule

20 tissues. The NFR polypeptides, which are essential for Nod-factor detection, comprise the first step in the Nod-factor signalling pathway that triggers the development of an infection thread and root nodules.

Nod-factor binding properties: are a characteristic of NFR1 and NFR5 polypeptides and are particularly associated with the extracellular domain of

25 said NFR polypeptides, which comprise LysM domains. The binding of Nod-factors by the extracellular domain of NFR polypeptides is specific, such the NFR polypeptides can distinguish between the strain-specific chemically modified forms of Nod-factor.

Nodulating plant: a plant capable of establishing an endosymbiotic

30 *Rhizobium* – plant interaction with a nitrogen-fixing *Rhizobium* bacterium, including the formation of an infection thread, and the development of root

nodules capable of fixing nitrogen. Nodulating plants are limited to a few plant families, and are particularly found in the *Legume* family, and they are all member of the Rosid 1 clade.

Non-nodulating plant: a plant which is incapable of establishing an

5 **endosymbiotic Rhizobia** – plant interaction with a nitrogen-fixing *Rhizobia* bacterium, and which does not form root nodules capable of fixing nitrogen.

Operably linked: refers to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

10 **ORF:** Open Reading Frame, which defines one of three putative protein coding sequences in a DNA polynucleotide.

Orthologue: Two homologous genes (or proteins) diverging concurrently with the organism harbouring them diverged. Orthologues commonly serve the same function within the organisms and are most often present in a similar position on the genome.

15 **PCR:** Polymerase Chain Reaction is a technique for the amplification of DNA polynucleotides, employing a heat stable DNA polymerase and short oligonucleotide primers, which hybridise to the DNA polynucleotide template in a sequence specific manner and provide the primer for 5' to 3' DNA

20 synthesis. Sequential heating and cooling cycles allow denaturation of the double-stranded DNA template and sequence-specific annealing of the primers, prior to each round of DNA synthesis. PCR is used to amplify DNA polynucleotides employing the following standard protocol or modifications thereof:

25 PCR amplification is performed in 25 μ l reactions containing: 10 mM Tris-HCl, pH 8.3 at 25°C; 50 mM KCl; 1.5 mM MgCl₂; 0.01% gelatin; 0.5 unit Taq polymerase and 2.5 pmol of each primer together with template genomic DNA (50-100 ng) or cDNA. PCR cycling conditions comprise heating to 94°C for 45 seconds, followed by 35 cycles of 94°C for 20 seconds; annealing at

30 X°C for 20 seconds (where X is a temperature between 40 and 70°C defined by the primer annealing temperature); 72°C for 30 seconds to several

minutes (depending on the expected length of the amplification product). The last cycle is followed by heating to 72°C for 2-3 minutes, and terminated by incubation at 4°C.

Pfam consensus: a consensus sequence derived from a large collection of 5 protein multiple sequence alignments and profile hidden Markov models used to identify conserved protein domains (Bateman *et al.*, 2002, Nucleic Acids Res. 30: 276-80; and searchable on <http://www.sanger.ac.uk/Software/Pfam/> and on NCBI at <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)

Protein domain prediction: sequences are analysed by BLAST 10 (www.ncbi.nlm.nih.gov/BLAST/) and PredictProtein (www.embl-heidelberg.de/predictprotein/predictprotein.html). Signal peptides are predicted by SignalP v. 1.1 (www.cbs.dtu.dk/services/SignalP/) and transmembrane regions are predicted by TMHMM v. 2.0 (www.cbs.dtu.dk/services/TMHMM/)

15 Polymorphism: refers to "DNA polymorphism" due to nucleotide sequence differences between aligned regions of two nucleic acid sequences.

Polynucleotide molecule: or "polynucleotide", or "polynucleotide sequence" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term 20 encompasses nucleic acids containing known analogs of natural nucleotides, which have similar binding properties as the reference nucleic acid.

Promoter: is an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary 25 nucleic acid sequences near the start site of transcription, e.g. a TATA box element, and optionally includes distal enhancer or repressor elements, which can be located several 1000bp upstream of the transcription start site. A tissue specific promoter is one which specifically regulates expressed in a particular cell type or tissue e.g. roots. A "constitutive" promoter is one that is 30 active under most environmental and developmental conditions throughout the plant.

RACE/5'RACE/3'RACE: Rapid Amplification of cDNA Ends is a PCR-based technique for the amplification of 5' or 3' regions of selected cDNA sequences which facilitates the generation of full-length cDNAs from mRNA.

The technique is performed using the following standard protocol or

5 modifications thereof: mRNA is reverse transcribed with RNase H⁻ Reverse Transcriptase essentially according to the protocol of Matz et al, (1999) *Nucleic Acids Research* 27: 1558-60 and amplified by PCR essentially according to the protocol of Kellogg et al (1994) *Biotechniques* 16(6): 1134-7.

10 Real-time PCR: a PCR-based technique for the quantitative analysis of mRNA species in an RNA preparation. The formation of amplified DNA products during PCR cycling is monitored in real-time, using a specific fluorescent DNA binding-dye and measuring fluorescence emission.

15 **Sexual cross:** refers to the pollination of one plant by another, leading to the fusion of gametes and the production of seed.

20 **SMART consensus:** represents the consensus sequence of a particular protein domain predicted by the Simple Modular Architecture Research Tool database (Schultz, J. et al. (1998)- *PNAS* 26;95(11):5857-64)

25 **Southern hybridisation:** Filters carrying nucleic acids (DNA or RNA) are prehybridized for 1-2 hours at 65°C with agitation in a buffer containing 7 % SDS, 0.26 M Na₂HPO₄, 5 % dextrane-suphate, 1 % BSA and 10µg/ml denatured salmon sperm DNA. Then the denatured, radioactively labelled

30 DNA probe is added to the buffer and hybridization is carried out over night at 65°C with agitation. For low stringency, washing is carried out at 65°C with a buffer containing 2XSSC, 0.1 % SDS for 20 minutes. For medium stringency, washing is continued at 65°C with a buffer containing 1XSSC, 0.1 % SDS for 2x 20 minutes and for high stringency filters are washed a further 2x 20 minutes at 65°C in a buffer containing 0.3XSSC, 0.1 % SDS.

35 Probe labelling by random priming is performed essentially according to Feinberg and Vogelstein (1983) *Anal. Biochem.* 132(1), 6-13 and Feinberg and Vogelstein (1984) Addendum. *Anal. Biochem.* 137(1), 266-267

Substantially identical: refers to two nucleic acid or polypeptide sequences that have at least 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the sequence

- 5 comparison algorithms given herein, or by manual alignment and visual inspection. This definition also refers to the complement of the test sequence with respect to its substantial identity to a reference sequence. A comparison window refers to any one of the number of contiguous positions in a sequence (being anything from between about 20 to about 600, most commonly about 100 to about 150) which may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Optimal alignment can be achieved using computerized implementations of alignment algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis. USA) or BLAST analyses available on the site: (www.ncbi.nlm.nih.gov/)
- 10
- 15

TAC clones: clones from a Transformation-competent Artificial Chromosome library.

- 20 **TM marker:** is a microsatellite marker developed from a TAC sequence, based on sequence differences between *Lotus japonicus* Gifu and MG-20 genotypes.

Transgene: refers to a polynucleotide sequence, for example a "transgenic expression cassette", which is integrated into the genome of a plant by means other than a sexual cross, commonly referred to as transformation, to give a transgenic plant.

- 25 **UTR:** untranslated region of an mRNA or cDNA sequence.
- 30 **Variant:** refers to "variant NFR1 or NFR5 polypeptides" encoded by different *NFR* alleles.
- Wild type:** a plant gene, genotype, or phenotype predominating in the wild population or in the germplasm used as standard laboratory stock.

II. Nod-factor binding

The present invention provides a Nod-factor binding element comprising one or more isolated NFR polypeptides. The isolated NFR polypeptides, NFR1 and NFR2, as exemplified by SEQ ID No: 25 and SEQ ID No: 8, bind to Nod-factors in a chemically-specific manner, distinguishing between the different chemically modified forms of Nod-factors produced by different *Rhizobium* strains. The chemical specificity of Nod-factor binding by NFR1 and NFR5 polypeptides is located in their extracellular domain, which comprises LysM type motifs. The LysM protein motif, first identified in bacterial lysis and

5 muramidase enzymes degrading cell wall peptidoglycans, is widespread among prokaryotes and eukaryotes (Pontig et al. 1999, *J Mol Biol.* 289, 729-745; Bateman and Bycroft, 2000, *J Mol Biol.* 299, 1113-1119). In bacteria it is often found in proteins associated with bacterial cell walls or involved in pathogenesis and *in vivo* and *in vitro* studies of *Lactococcus lactis* autolysin

10

15 demonstrate that the three LysM domains of this protein bind peptidoglycan (Steen et al, 2003, *J Biol. Chem.* April issue). Since both A- and B-type peptidoglycans, differing in amino acid composition as well as cross-linking were bound, it was concluded that autolysin LysM domains binds the N-acetyl-glucosamine-N-acetyl-mureine backbone polymer LysM domains are

20 frequently found together with amidase, protease or chitinase motifs and two confirmed chitinases carry LysM domains. One is the sex pheromone and wound-induced polypeptide from the alga *Volvox carteri* that binds and degrades chitin *in vitro* (Amon et al. 1998, *Plant Cell* 10, 781-9). The other is α -toxin from *Kluyveromyces lactis*, that docks onto a yeast cell wall chitin

25 receptor (Butler, et al. (1991) *Eur J Biochem* 199, 483-8). Structure-based alignment of representative LysM domain sequences have shown a pronounced variability among their primary sequence, except the amino acids directly involved in maintaining the secondary structure.

The NFR polypeptides are transmembrane proteins, able to transduce

30 signals perceived by the extracellular NFR domain across the membrane to the intracellular NFR domain comprising kinase motifs, which serves to

couple signal perception to the common symbiotic signalling pathway leading to nodule development and nitrogen fixation.

The methods employed for the practise and understanding of the invention, which are described below, involve standard recombinant DNA technology
5 that are well-known and commonly employed in the art and available from Sambrook *et al.*, 1989, *Molecular Cloning: A laboratory manual*.

III. Isolation of nucleic acid molecules comprising *NFR* genes and cDNAs encoding *NFR1* and *NFR5* polypeptides and their orthologues.

10 The isolation of genes and cDNAs encoding *NFR1* and *NFR5* polypeptides, comprising an amino acid sequence substantially similar to SEQ ID No: 25 or SEQ ID No: 8 respectively, may be accomplished by a number of techniques. For instance, a BLAST search of a genomic or cDNA sequence bank of a desired legume plant species (e.g. soybean, pea or *Medicago truncatula*) can
15 identify test sequences similar to the *NFR1* or *NFR5* reference sequence, based on the smallest sum probability score (P(N)). The (P(N)) score (the probability of the match between the test and reference sequence occurring by chance) for a "similar sequence" will be less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.
20 This approach is exemplified by the *Medicago truncatula* sequence (Ac126779) included in Figure 3. Oligonucleotide primers, together with PCR, can be used to amplify regions of the test sequence from genomic or cDNA of the selected plant species, and a test sequence which is similar to the full-length *NFR1* and *NFR5* gene sequences can be assembled. In the
25 case that an appropriate gene bank is not available for the selected plant species, oligonucleotide primers, based on *NFR1* and *NFR5* gene sequences, can be used to PCR amplify similar sequences from genomic or cDNA prepared from the selected plant. Alternatively, nucleic acid probes based on *NFR1* and *NFR5* gene sequences can be hybridised to genomic or
30 cDNA libraries prepared from the selected plant species using standard conditions, in order to identify clones comprising sequences similar to *NFR1*

or *NFR5* genes. A nucleic acid sequence in a library, which hybridises to a *NFR1* or *NFR5* gene-specific probe under conditions which include at least one wash in 2xSSC at a temperature of at least about 65°C for 20 minutes, is potentially a similar sequence to a *NFR1* or *NFR5* gene. A test sequence

5 comprising a full-length cDNA sequence similar to *NFR1* or *NFR5* cDNAs having SEQ ID No: 22 and 22 or SEQ ID no: 6 respectively, can be generated by 5' RACE cDNA synthesis, as described herein. The nucleic acid sequence of each test sequence, derived from a selected plant species, is determined in order to identify nucleic acid molecules which

10 are substantially identical to *NFR1* or *NFR5* genes having SEQ ID No: 23 or SEQ ID No: 7Y respectively, or nucleic acid molecules that encode proteins whose amino acid sequence is substantially identical to *NFR1* or *NFR5*, having SEQ ID No: 25 or SEQ ID No. 8, respectively.

15 **IV. Transgenic plants expressing *NFR1* and/or *NFR5* polypeptides**
The polynucleotide molecules of the invention can be used to express a Nod-factor binding element in non-nodulating plants and thereby confer the ability to bind Nod-factors and establish a *Rhizobium*/plant interaction leading to nodule development. An expression cassette comprising a nucleic acid

20 sequence encoding a *NFR* polypeptide, substantially identical to SEQ ID No: 25 or SEQ ID No: 8, and operably linked to its own promoter or a heterologous promoter and 3' terminator can be transformed into a selected host plant using a number of known methods for plant transformation. By way of example, the expression cassette can be cloned between the T-DNA

25 borders of a binary vector, and transferred into an *Agrobacterium tumefaciens* host, and used to infect and transform a host plant. The expression cassette is commonly integrated into the host plant in parallel with a selectable marker gene giving resistance to an herbicide or antibiotic, in order to select transformed plant tissue. Stable integration of the expression

30 cassette into the host plant genome is mediated by the virulence functions of the *Agrobacterium* host. Binary vectors and *Agrobacterium tumefaciens*-

based methods for the stable integration of expression cassettes into all major cereal plants are known, as described for example for rice (Hiei *et al.*, 1994, *The Plant J.* 6: 271-282) and maize (Yuji *et al.*, 1996, *Nature Biotechnology*, 14: 745-750). Alternative transformation methods, based on

5 direct transfer can also be employed to stably integrate expression cassettes into the genome of a host plant, as described by Miki *et al.*, 1993, "Procedure for introducing foreign DNA into plants", In: *Methods in Plant Molecular Biology and Biotechnology*, Glick and Thompson, eds., CRC Press, Inc., Boca Raton, pp 67-88). Promoters to be used in the expression cassette of

10 the invention include constitutive promoters, as for example the 35S CaMV promoter ((Acc V00141 and J02048) or in the case of a cereal host plant the Ubi1 gene promoter (Christensen *et al.*, 1992, *Plant Mol Biol* 18: 675-689). In a preferred embodiment, a root specific promoter is used in the expression cassette, for example the maize zmGRP3 promoter (Goodemeir *et al.* 1998, *Plant Mol Biol*, 36, 799-802) or the epidermis expressed maize promoter described by Ponce *et al.* 2000, *Planta*, 211, 23-33. Terminators that may be used in the expression construct can for instance be the NOS terminator (Acc NC_003065).

15 Host plants transformed with an expression cassette encoding one NFR polypeptide, for example NFR1, or its orthologue, can be crossed with a second host plant transformed with an expression cassette encoding a second NFR polypeptide, for example NFR, or its orthologue. Progeny expressing both said NFR polypeptides can then be selected and used in the invention. Alternatively, host plants can be transformed with a vector

20 comprising two expression cassettes encoding both said NFR polypeptides.

25

V. NFR genes encoding NFR polypeptide having specific Nod-factor binding properties.

30 Nucleic acid molecules comprising *NFR1* or *NFR5* genes encoding NFR polypeptides having specific Nod-factor binding properties can be identified

by a number of functional assays described in the "Examples" given herein. In a preferred embodiment, said nucleic acid sequences are expressed transgenically in a host plant employing the expression cassettes described above. Expression of *NFR1* or *NFR5* genes or their homologous/orthologues

- 5 In plant roots allows the specific Nod-factor binding properties of the expressed *NFR* protein to be fully tested. Assays suitable for establishing specific Nod-factor binding include the detection of: a morphological root hair response (e.g. root hair deformation, root hair curling); a physiological response (e.g. root hair membrane depolarisation, ion fluxes, pH changes and calcium oscillations); a symbiotic signalling response (e.g. downstream activation of symbiotic nodulin gene expression) following root infection with *Rhizobium* bacteria or isolated Nod-factors; the ability to develop root nodule primordia, infection pockets or root nodules, where the response is strain dependent or dependent on the chemical modification of Nod-factor structure.
- 10
- 15

VI. Marker assisted breeding for *NFR* alleles.

A method for marker assisted breeding of *NFR* alleles, encoding variant *NFR* polypeptides, is described herein, with examples from *Lotus* *NFR* alleles. In summary, variant *NFR1* or *NFR5* polypeptides, comprising an amino acid sequence substantially similar to SEQ ID No: 25 or SEQ ID No: 8 respectively, are identified in a nodulating legume species, and the *Rhizobium* strain specificity of said variant *NFR1* or *NFR5* polypeptide is determined, according to measurable morphological or physiological parameters described herein. Subsequently, DNA polymorphisms at loci genetically linked to, or within, the gene locus encoding said variant *NFR1* or *NFR5* polypeptide, are identified on the basis of the nucleic acid sequence of the loci or its neighbouring DNA region. Molecular markers based on said DNA polymorphisms, are used for the identification and selection of plants carrying *NFR* alleles encoding said variant *NFR1* or *NFR5* polypeptides. Use

- 20
- 25
- 30

of this method provides a powerful tool for the breeding of legumes with enhanced nodulation frequency.

III. Examples

5 Example 1.

Cloning of Nod-factor Binding Element Genes

Genetic studies in the legume plants *Lotus japonicus* (*Lj*) and pea (*Ps*) have generated collections of symbiotic mutants, which have been screened for mutants blocked in the early steps of symbiosis (Geurts and Bisseling, 2002
10 *supra*; Kistner and Parniske 2002 *Trends in Plant Science* 7: 511-518). Characteristic for a group of the selected mutants is their inability to respond to Nod-factors, with the absence of root hair deformation and curling, cortical cell division to form the cortical primordium, and induction of the early nodulin genes which contribute to nodule development and function. Nod-factor
15 induced calcium oscillations were also found to be absent in some of these mutants, indicating that they are blocked in an early step in Nod-factor signalling. Among this latter group, are a few mutants, including members of the *Pssym10* complementation group and *LjNFR1* and *LjNFR5* (previously called *Ljsym1 and 5*), which failed to respond to Nod-factors but retain their
20 ability to establish mycorrhizal associations. Genetic mapping indicates that pea *SYM10* and *Lotus NFR5* loci in the pea and *Lotus* could be orthologs. Mutants falling within this group provided a useful starting point in the search for genes encoding potential candidate proteins involved in Nod-factor binding and perception.

25 **A. Isolation, cloning and characterisation of *NFR5* genes and gene products.**

1. Map based cloning of *Lj NFR5*

The symbiotic mutants of *Lotus japonicus* *nfr5-1*, *nfr5-2* and *nfr5-3* (also known as *sym5*), (previously isolated by Schäuser et al 1998 *Mo. Gen Genet*, 259: 414-423; Szczęśniowski et al 1998, *Mol Plant-Microbe Interact*, 11: 684-697) were utilised. To determine the root nodulation phenotype under 5 symbiotic conditions, seeds were surface sterilised in 2% hyperchlorite, washed and inoculated with a two day old culture of *M. loti* NZP2235. Plants were cultivated in the nitrogen-free B&D nutrients and scored after 6-7 weeks (Broughton and Dilworth, *Biochem J*, 1971, 125, 1075-1080; Handberg and Stougaard, *Plant J*. 1992, 2, 487-496). Under non-symbiotic conditions, plants 10 were cultivated in Hormum nutrients (Handberg and Stougaard, *Plant J*. 1992, 2, 487-496).

Mapping populations were established in order to localise the *nfr5* locus on the *Lotus japonicus* genome. Both intra- and interspecific F2 mapping populations were created by crossing a *Lotus japonicus* "Gifu" *nfr5-1* mutant 15 to wild type *Lotus japonicus* ecotype "MG20" and to wild type *Lotus filicaulis*. MG-20 seeds are obtainable from Sachiko ISOBE, National Agricultural Research Center for Hokkaido Region, Hitsujigaoka, Toyohira, Sapporo Hokkaido 062-8555, JAPAN and *L. filicaulis* from Jens Stougaard, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, 20 DK-8000 Aarhus C. F2 plants homozygous for the *nfr5-1* mutant allele were identified after screening for the non-nodulation mutant phenotype. 240 homozygous F2 mutant plants were analysed in the *L. filicaulis* mapping population and 368 homozygous F2 mutant plants in the "MG20" mapping population. 25 Positional cloning of the *nfr5* locus was performed by AFLP and Bulked Segregant Analysis of the mapping populations using the *Eco*RI/*Mse*I restriction enzyme combination (Vos et al, 1995, *Nucleic Acids Res*.23, 4407-4414; Sandal et al 2002, *Genetics*, 161, 1673-1683). Initially, *nfr5* was mapped to the lower arm of chromosome 2 between AFLP markers E33M40-30 E22F and E32M54-12F in the *L. filicaulis* based mapping population, as

shown in Figure 1a. The E32M54-12F marker was cloned and used to isolate BAC clones BAC8H12 and BAC67I22 and TAC clone LjT18J10, as shown in Figure 1b. The ends of this contig were used to isolate adjacent BAC and TAC clones namely BAC58K7 and LjT01C03 at one end and TAC 5 LjB06D23 on the other end. The outer end of LjB06D23 was used to isolate TAC clone LjT13I23. The outer end of LjB06D23 was used to isolate TAC clone LjT13I23 (TM0522). Various markers from this contig were mapped on the mapping populations from *nfr5-1* crossed to *L. filicaulis* and to *L. japonicus* MG-20. In the *L. filicaulis* mapping population one recombinant 10 plant was found with the outer end of the TAC clone TM0522, whereas no recombinant plants were found with a marker from the middle of this TAC clone. In the *L. japonicus* MG-20 mapping population, 4 recombinant plants out of 368 plants were found with the marker TM0323, thereby delimiting *nfr5* to a region of 150 kb. This region was sequenced and found to contain 13 15 ORFs, of which two encoded putative proteins sharing sequence homology to receptor kinases. Sequencing of these two specific ORFs in genomic DNA derived from *nfr5-1* showed that one of the ORF sequences contained a 27 nucleotide deletion. Furthermore sequencing of this ORF in genomic DNA from *nfr5-2* and *nfr5-3* showed the insertion of a retrotransposon and a point 20 mutation leading to a premature stop codon, respectively, as shown in Figure 1d. The localisation of the *nfr5* locus from physical and genetic mapping data, combined with the identification of mutations in three independent *nfr5* mutant alleles, provides unequivocal evidence that mutations in the *NFR5* ORF lead to a loss of Nod-factor perception.

25 **2. Cloning the *Lj NFR5* cDNA**

A full-length cDNA corresponding to the *NFR5* gene was isolated using a combination of 5' and 3' RACE. RNA was extracted from *Lotus japonicus* roots, grown in the absence of nitrate or rhizobia, and reverse transcribed to make a full-length cDNA pool for the performance of 5'-RACE according to 30 the standard protocol. The cDNA was amplified using the 5' oligonucleotide

5'CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT 3'
(SEQ ID No:1) and the reverse primer
5'GCTAGTTAAAATGTAATAGTAACCACGC3' (SEQ ID No: 2), and a
RACE-product of approximately 2 kb was cloned into a topoisomerase
5 activated plasmid vector (Shuman, 1994, *J Biol Chem* 269: 32678-32684). 3'-
RACE was performed on the same 5'-RACE cDNA pool, using a 5' gene-
specific primer 5' AAAGCAGCATTCATCTTCTGG 3' (SEQ ID No: 3) and an
oligo-dT primer 5'GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTV
3' (SEQ ID No: 4), where the first 5 PCR cycles were carried out at an
10 annealing temperature of 42° C and the following 30 cycles at higher
annealing temperature of 58° C. The products of this PCR reaction were used
as template for a second PCR reaction with a gene-specific primer positioned
further 3' having the sequence 5' GCAAGGGAAGGTAATTCAG 3' (SEQ ID
No: 5) and the above oligo dT-primer, using standard PCR amplification
15 conditions (annealing at 54° C; extension 72° C for 30 s) and the products
cloned into a topoisomerase activated plasmid vector (Shuman, 1994,
supra). Nucleotide sequencing of 18 5'RACE clones and three 3' RACE
clones allowed the full-length sequence of the *NFR5* cDNA to be determined
(SEQ ID No: 6). The *NFR5* cDNA was 2283 nucleotides in length, with an
20 open reading frame of 1785 nucleotides, preceded by a 5' UTR leader
sequence of 140 nucleotides and a 3'UTR region of 358 nucleotides.
Alignment of the *NFR5* cDNA sequence with the *NFR5* gene sequence
(SEQ ID No: 7), shown schematically in Figure 1d, confirmed that the gene is
devoid of introns.
25
3. Primary sequence and structural domains of LjNFR5 and mutant alleles.
The primary sequence and domain structure of NFR5, encoded by *NFR5*, are
consistent with a transmembrane Nod-factor binding protein, required for
30 Nod-factor perception in rhizobial-legume symbiosis. The *NFR5* gene
encodes an NFR5 protein of 596 amino acids having the sequence given in

Figure 2b (SEQ ID No: 8) and a predicted molecular mass of 65.3 kD. The protein domain structure predicted for NFR5 and shown in Figure 2a,b, defines a signal peptide, comprising a hydrophobic stretch of 26 amino acids, followed by an extracellular domain with three LysM-type motifs, a

5 transmembrane domain and an intracellular kinase domain. The LysM-type motifs found in *Lotus* NFR5, SYM10, *Medicago truncatula* (M.t, Ac126779), and by homology in a rice gene (Ac103891), show homology to the single LysM motif present in an algal (*Volvox carteria*) chitinase (Amon *et al*, 1998, *Plant Cell* 10: 781-789) and to the Pfam consensus, as illustrated in the

10 amino acid sequence alignment of this domain given in Figure 2c. The NFR5 kinase domain has motifs characteristic of functional serine/threonine kinases (Schenk and Snaar-Jagalska, 1999, *Biochim Biophys Acta* 1449: 1-24; Huse and Kuriyan, 2002, *Cell* 109: 275-282), with the exception that motif VII lacks an aspartic acid residue conserved in kinases, and motif VIII, comprising the

15 activation loop, is either divergent or absent.

Analysis of the *nfr5* mutant genes reveals that the point mutation in *nfr5-3* and the retrotransposon insertion in *nfr5-2* will express truncated polypeptides of 54 amino acids, lacking the LysM motifs and entire kinase domain; or of 233 amino acids, lacking the kinase motifs X and XI,

20 respectively. The 27 nucleotide deletion in the *nfr5-1* mutant removes 9 amino acids from kinase motif V.

4. Cloning and characterisation of the pea SYM10 gene and cDNA and sym10 mutants.

25 Wild type pea cv's (Alaska, Finale, Frisson, Sparkle) and the symbiotic mutants (N15; P5; P56) were obtained from the pea germ-plasm collection at JIC Norwich-UK, while the symbiotic mutant, RisFixG, was obtained from Kjeld Engvild, Risø National Laboratory, 8000 Roskilde, Denmark. The mutants, belonging to the pea *sym10* complementation group, were identified

30 in the following genetic backgrounds: N15 type strain in a Sparkle

background (Kneen *et al*, 1994, *J Heredity* 85: 129-133), P5 in a Frisson background (Duc and Messager, 1989, *Plant Science* 60: 207-213), RisFixG in a Finale background RisFixG (Engvild,1987, *Theoretical Applied Genetics* 74: 711-713; Borisov *et al.*, 2000, *Czech Journal Genetics and Plant Breeding* 36: 106-110); P56 in a Frisson background (Sagan *et al*.1994, *Plant Science* 100: 59-70).

A fragment of the pea *SYM10* gene was cloned by PCR amplification of cv Finale genomic DNA using a standard PCR cycling program and the forward primer 5'-ATGTCTGCCTCTTCCTTC-3', (SEQ ID No: 9) and the reverse primer 5'-CCACACATAAGTAATMAGATACT-3', (SEQ ID No: 10). The sequence of these oligonucleotide primers was based on nucleotide sequence stretches conserved in *L. japonicus* *NFR5* and the partial sequence of an *NFR5* homologue identified in a *M. truncatula* root EST collection (BE204912). The identity of the amplified 551 base pair *SYM10* product was confirmed by sequencing, and then used as a probe to isolate and sequence a pea cv Alaska *SYM10* genomic clone (SEQ ID No:11) from a cv. Alaska genomic library (obtained from H. Franssen, Department of Molecular Biology, Agricultural University, 6703 HA Wageningen, The Netherlands) and a full-length pea cv. Finale *SYM10* cDNA clone (SEQ ID No: 12) from a cv. Finale cDNA library (obtained from H. Franssen, *supra*), which were then sequenced. The sequence of the *SYM10* gene in cv. Frisson (SEQ ID No:13) and in cv. Sparkle (SEQ ID No: 14) were determined by a PCR amplification and sequencing of the amplified gene fragment. The nucleotide sequence of the corresponding mutants P5, P56, and RisFixG were also determined by a PCR amplification and sequencing of the amplified gene fragment.

Nucleotide sequence comparison of the *SYM10* gene in the *Pssym10* mutant lines (P5, RisFix6 and P56) with the wild type parent lines revealed, in each case, sequence mutations, which could be correlated with the mutant

phenotype. The 3 independent *sym10* mutant lines identified 3 mutant alleles of the *SYM10* gene, all carrying nonsense mutations, and the N15 type strain was deleted for *SYM10* (Table 1, Figure 4c). Southern hybridization with probes covering either the extracellular domain of *SYM10* 5 or the 3'UTR on *EcoRI* digested DNA from N15 and the parent variety Sparkle, shows that the *SYM10* gene is absent from the N15 mutant line.

5. Primary sequence and structural domains of *PsSYM10* and mutant alleles.

10 The *PsSYM10* protein of pea, encoded by *PsSYM10*, is a homologue of the NFR5 transmembrane Nod-factor binding protein of *Lotus*, required for Nod-factor perception in rhizobial-legume symbiosis. The pea cv Alaska *SYM10* gene encodes a *SYM10* protein (SEQ ID No: 15) of 594 amino acid residues, with a predicted molecular mass of 66 kD, which shares 73% amino acid 15 identity with the NFR5 protein from *Lotus*. In common with the NFR5 protein, the *SYM10* protein has an N-terminal signal peptide, an extracellular region with three LysM motifs, followed by a transmembrane domain, and then an intracellular domain comprising kinase motifs (Figure 2 and 3). The *sym10* genes in the symbiotic pea mutants P5, RisFix6 and P56, each 20 having premature stop codons, encode truncated *SYM10* proteins of 199, 387 and 404 amino acids, respectively, which lack part of, or the entire, kinase domain (Table 1).

6. The NFR5 protein family is unique to nodulating plants

25 Comparative analysis defines *LjNFR5* and *PsSYM10* as members of a novel family of transmembrane Nod-factor binding proteins. A BLAST search of plant gene sequences suggests that genes encoding related, but presently uncharacterised, proteins may be present in the legume *Medicago truncatula* (Ac126779), while more distantly related, predicted proteins may be found in 30 rice (Ac103891) and *Arabidopsis* (At2g33580), with a sequence identity to NFR5 of 61%, 39%, and 28%, respectively. The high level of sequence

conservation in *M. truncatula* (Ac126779) makes this protein and the gene encoding the protein substantially identical to NFR5. In common with the NFR5 and SYM10, the kinase domains of these proteins also lack the conserved aspartic acid residue of motif VII, and the activation loop in motif

5 VIII is highly diverged or absent, as shown in Figure 2d, with the exception of the *Arabidopsis* protein. Only distantly related proteins are therefore found outside the legume family. In conclusion, the NFR5 protein family appears to be restricted to nodulating legumes, and its absence from other plant families may be a key limiting factor in the establishment of rhizobial-root interactions

10 in the members of the families.

7. Tissue specific expression of the *LjNFR5* and *PsSYM10* genes

The expression pattern of the *NFR5* and *SYM10* genes in *Lotus* and pea is consistent with the role of their gene products as transmembrane Nod-factor binding proteins in the perception of rhizobial Nod-factors at the root surface and later during tissue invasion.

The expression of the *NFR5* and *SYM10* genes in various isolated organs of *Lotus* and pea plants, was investigated by determining the steady state *NFR5* and *SYM10* mRNA levels using Real-time PCR and/or Northern blot analysis. Total RNA was isolated from root, leaf, flower, pod and nodule tissues of uninoculated or inoculated *Lotus* "Gifu" or pea plants using a high salt extraction buffer followed by purification through a CsCl cushion. For Northern analysis, according to standard protocols, 20 µg total RNA was size-fractionated on 1.2% agarose gel, transferred to a Hybond membrane, hybridised overnight with an *NFR5* or *SYM10* specific probe covering the extracellular domain and washed at high stringency. Hybridization to the constitutively expressed ubiquitin *UBI* gene was used as control for RNA loading and quality of the RNA.

For the quantitative real-time RT-PCR, total RNA was extracted using the CsCl method and the mRNA was purified by biomagnetic affinity separation

(Jakobsen, K.S. et al (1990) *Nucleic Acids Research* 18(12): 3669). The RNA preparations were analysed for contaminating DNA by quantitative PCR and when necessary, the RNA was treated with DNaseI. The DNaseI enzyme was then removed by phenol:chloroform extraction and the RNA was 5 precipitated and re-suspended in 20 µl RNase free H₂O. First strand cDNA was prepared using Expand reverse transcriptase and the quantitative real-time PCR was performed on a standard PCR LightCycler instrument. The efficiency-corrected relative transcript concentration was determined and normalized to a calibrator sample, using *Lotus japonicus* ATP synthase gene 10 as a reference (Gerard C.J. et al, 2000 *Mol. Diagnosis* 5: 39-45).

The level of *NFR5* mRNA, determined by Northern blot analysis and quantitative RT-PCR, was 60 to 120 fold higher in the root tissue of *Lotus* plants in comparison to other plant tissues (leaves, stems, flowers, pods, and nodules), as shown in Figure 4a. Northern hybridisation show highest 15 expression of *NFR5* in *Lotus* root tissue and a barely detectable expression in nodules. Northern blot analysis detected *SYM10* mRNA in the roots of pea, and a higher level in nodules, but no mRNA was detected in leaves, as shown in Figure 4c.

20 **B. Isolation, cloning and characterisation of *NFR1* genes and gene products.**

1. Map based cloning of *Lj NFR1*

The *NFR1* gene was isolated using a positional cloning approach. On the genetic map of *Lotus* the *NFR1* locus is located on the short arm of chromosome I, approximately 22 cM from the top, within a 7.6 cM interval, as 25 shown in Figure 5a. Several TM markers and PCR markers, derived from DNA polymorphism in the genome sequences of the *L. japonicus* mapping parents, were found to be closely linked to *NFR1* locus and were used to narrow down the region. A physical map of the region, comprising a contig of

assembled BAC and TAC clones, is shown in figure 5b. Fine mapping in an F2 population, established from a *Lotus japonicus* *nfr-1* mutant to wild type *L. japonicus* ecotype 'Miyakojima MG-20' cross, and genotyping of 1603 mutant plants, identified two markers (56K22, 56L2-2) delimiting the *NFR1* locus

5 within a region of 250 kb. BAC and TAC libraries, available from Satoshi Tabata, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0812 Japan; another BAC library from Jens Stougaard, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, DK-8000 Aarhus C, were screened using the closest flanking markers (56L2-1, 10M24-1, 36D15) as probes, and the *NFR1* locus was localised to 36 kb within the region. The

10 ORFs detected within the region coded for a UFD1-like protein, a hypothetical protein and a candidate *NFR1* protein showing homology to receptor kinases, (Figure 5b).

The region in the genomes of *nfr1-1*, *nfr1-2* mutants, corresponding to the

15 candidate *NFR1* gene was amplified as three fragments by PCR under standard conditions and sequenced. The fragment of 1827 bp amplified using PCR forward primer 5'TGC ATT TGC ATG GAG AAC C3', (SEQ ID No: 16) and reverse primer 5' TTT GCT GTG ACA TTA TCA GC3', (SEQ ID No: 17) contains single nucleotide substitutions leading to translational stop codons

20 in both the mutant alleles *nfr1-1*, with a CAA to TAA substitution, and the *nfr1-2*, with a GAA to TAA substitution. The physical and genetic mapping of the *nfr1* locus, combined with the identification of mutations in two independent *nfr1* mutant alleles, provides unequivocal evidence that the sequenced *NFR1* gene is required for Nod-factor perception and subsequent

25 signal transduction.

2. Cloning the *Lj* *NFR1* cDNAs

Two alternatively spliced *Lj* *NFR1* cDNAs were identified using a combination of cDNA library screening and 5' RACE on root RNA from *Lotus japonicus*. A *Lotus* root cDNA library (Poulsen *et al.*, 2002, *MPMI* 15:376-379) was

30 screened with an *NFR1* gene probe generated by PCR amplification of the

nucleotides between 9689 to 10055 of the genomic sequence, using the primer pair: 5' TTGCAGATTGCACAACTAGG3' (SEQ ID No: 18) and 5'ACTTAGAATCTGCAACTTGC 3' (SEQ ID No: 19). Total RNA extracted from *Lotus* roots, was amplified by 5' RACE, according to the standard

5 protocol, using the gene specific reverse primer 5'ACTTAGAATCTGCAACTTGC 3' (SEQ ID No 20). Based on the sequence of isolated *NFR1* cDNAs and 5' RACE products, the *NFR1* gene produces two mRNA species, of 2187 (SEQ ID No: 21) and 2193 nucleotides (SEQ ID No: 22), with a 5' leader sequence of 114 nucleotides, and a 3'

10 untranslated region is 207 nucleotides (Figure 5c). Alignment of genomic and cDNA sequences defined 12 exons in *NFR1* and a gene structure spanning 10235 bp (SEQ ID No: 23). The sequenced region includes 4057bp from the stop codon of the previous gene up to the transcription start point of *NFR1* + 6009 bp of *NFR1* + 187 bp of 3' genomic. Alternative splice donor sites at the

15 3' of exon IV account for the two alternative *NFR1* mRNA species.

3. Primary sequence and structural domains of *LjNFR1* and mutant alleles.

The primary sequence and domain structure of *NFR1*, encoded by *LjNFR1*, 20 are consistent with a transmembrane Nod-factor binding protein, required for Nod-factor perception in *Rhizobium*-legume symbiosis. The alternatively spliced *NFR1* cDNAs encode *NFR1* proteins of 621 (SEQ ID No: 24) and 623 amino acids (SEQ ID No: 25), with a predicted molecular mass of 68.09 kd and 68.23 kd, respectively. The protein has an amino-terminal signal peptide, 25 followed by an extracellular domain having two LysM-type motifs, a transmembrane domain, and an intracellular carboxy-terminal domain comprising serine/threonine kinases motifs. In *nfr1-1*, a stop codon in kinase domain VIII encodes truncated polypeptides of 490 and 492 amino acids, and in *nfr1-2* a stop codon between domain IX 30 and XI encodes truncated polypeptides of 526 and 528 amino acids, as indicated in Figure 6a.

In Figure 6b the M1 LysM motif of NFR1 is aligned with the LysM motifs from *Arabidopsis thaliana* and the SMART consensus and M2 LysM of NFR1 with the *Volvox carteri* chitinase (Acc No: T08150), the closest related *Arabidopsis thaliana* receptor kinase (Acc No: NP_566689) , the rice (Acc No:

5 BAB89226) and the consensus SMART LysM motif.

4. The LjNFR1 protein family is not found in non-nodulating plants

Comparative analysis defines LjNFR1 as a member of a second novel family of transmembrane Nod-factor binding proteins. Although proteins having both

10 receptor-like kinase domains and LysM motifs are predicted from plant genome sequences, their homology to NFR1 is low and their putative function unknown. *Arabidopsis* has five predicted receptor-like kinases with LysM motifs in the extracellular domain, and one of them (At3g21630) is 54% identical to NFR1 at the protein level. Rice has 2 genes in the same class, 15 and one (BAB89226) encodes a protein with 32 % identity to NFR1. This suggests that the NFR1 protein is essential for Nod-factor perception and its absence from non-nodulating plants may be a key limiting factor in the establishment of rhizobial-root interactions in these plants. Although NFR1 shares the same domain structure to NFR5 their primary sequence 20 homology is low (Figure 11).

5. Expression of the LjNFR1, NFR5 and SymRK symbiotic genes is root specific and independently regulated.

The NFR1 dependent root hair curling, in the susceptible zone located just 25 behind the root tip, is correlated with root specific NFR1 gene expression. Steady-state NFR1 mRNA levels were measured in different plant organs using quantitative real-time PCR and Northern blot analysis as described above in section A.7. NFR1 mRNA was only expressed in root tissue, and remained below detectable levels in leaves, flowers, pods and nodules, as 30 shown in Figure 7a. Upon inoculation with *M. loti*, the expression of NFR1 in wild type plants is relatively stable for at least 12 days after inoculation

(Figure 7b). Real-time PCR experiments revealed no difference between the levels of the two *NFR1* transcripts detected in the root RNA, suggesting that the alternative splicing of exon 4 is not differentially regulated.

5 *NFR1*, *NFR5* and *SymRK* gene expression in roots, before and following *Rhizobium* inoculation, was determined by real-time PCR in wild type and *nfr1*, *nfr5* and *symrk* mutant genotypes. The expression of *NFR1*, *NFR5* and *SymRK* genes in un-inoculated and inoculated roots was not significantly influenced by the symbiotic mutant genotype (Figure 7b, c, d) indicating that transcriptional regulation of these genes is mutually independent.

10

Example 2.

Functional properties of the Nod-factor binding element and its component NFR proteins

15 The functional and regulatory properties of the Nod-factor binding element and its component NFR proteins provide valuable tools for monitoring the functional expression and specific activity of the NFR proteins. Nod-factor perception by the Nod-factor binding element triggers the rhizobial-host interaction, which includes depolarisation of the plasma membrane, ion fluxes, alkalization of the external root hair space of the invasion zone,

20 calcium oscillations and cytoplasmic alkalization in epidermal cells, root hair morphological changes, infection thread formation and the initiation of the nodule primordia. These physiological events are accompanied and coordinated by the induction of specific plant symbiotic genes, called nodulins. For example, the *NIN* gene encodes a putative transcriptional

25 regulator facilitating infection thread formation and inception of the nodule primordia and limits the region of root cell-rhizobial interaction competence to a narrow invasion zone (Geurts and Bisseling, 2002, *supra*). Since *nin* mutants develop normal mycorrhiza, the *NIN* gene lies in the rhizobia-specific branch of the symbiotic signalling pathway, downstream of the common

30 pathway. Ion fluxes, pH changes, root hair deformation and nodule formation

are all absent in *NFR1* and *NFR5* mutant plants, and hence the functional activity of these genes must be required for all downstream physiological responses. Several physiological and molecular markers that are diagnostic of *NFR* expression are provided below.

5

1. Morphological marker of *NFR1* and *NFR5* gene expression

When wild type *Lotus japonicus* plants are inoculated with *Mesorhizobium loti*, the earliest visible evidence of infection is root hair deformation and root hair curling, which occurs 24 hours after inoculation, as shown in Figure 8a.

10 However, mutant plants carrying the *nfr1-1* (Figure 8c), *nfr1-2*, *nfr5-1*, *nfr5-2* or *nfr5-3* alleles (as in Figure 8c), all failed to produce root hair curling or deformation, infection threads or nodule primordia in response to infection by *Mesorhizobium loti* with all three strains tested (NZP2235, R7A and TONO). Lipochitin-oligosaccharides purified from *M. loti*, R7A strain, which induce

15 root hair deformation and branching in wild type plants (Figure 8b), also failed to induce any deformation of root hairs of the *nfr1-1* and *nfr5-1* mutants (Figure 8d), evidencing the key role of the *NFR1* and *NFR5* genes in Nod-factor perception.

20 Mutations in genes expressing the downstream components of the symbiosis signalling pathway, namely *symRK* and *nin* have clearly distinguishable phenotypes. After infection with *Mesorhizobium loti*, the root hairs of *symRK* plants swell into balloon structures (Figure 8e), while the *nin* mutants produce an excessive root hair response (Figure 8g). The response of double mutants carrying *nfr1-1/symRK-3* mutant alleles or *nfr1-1/nin* alleles to

25 *Mesorhizobium loti* infection (Figure 8f,h) are similar to that of *nfr1-1* mutants, demonstrating that the *nfr1-1* mutation is dominant to *symRK* and *nin* mutations, and hence determines an earlier step in the symbiotic signalling pathway.

30 2. Physiological marker of *NFR1* and *NFR5* gene expression

When the root hairs of wild type *Lotus* plants are exposed to *M. loti* Nod-factor, the plasma membrane is depolarised and an alkalinisation occurs in the root hair space of the invasion zone, (Figure 9a). The extracellular pH was monitored continuously in a flow-through regime using a pH-selective

5 microelectrode, placed within the root hair space. Membrane potential was measured simultaneously with pH, and the calculated values are based on at least three equivalent experiments, each. Mutants carrying *nfr1* and *nfr5* alleles do not respond normally to Nod-factor stimulation. Two *nfr5* alleles abolish the response to Nod-factors (Figure 9b), while the *nfr1-1* allele

10 causes a diminished and slower alkalinisation, and the *nfr1-2* allele causes the acidification of the extracellular root hair space (Figure 9c). Both the *NFR1* and *NFR5* genes are thus essential for mounting the earliest detectable cellular and electrophysiological responses to Nod-factor, which can be used to monitor their functional activity.

15 The early physiological response of the *symRK-3* and *symRK-1* mutant plants to *Mesorhizobium loti* Nod-factor is similar to the wild type (Figure 9d) and clearly distinguishable from the response of both the *nfr1* and *nfr5* mutants.

The response of the double mutant, carrying *nfr1-2/symRK-3* mutant alleles,

20 to Nod-factor (Figure 9e) is similar to that of *nfr1-2* mutants, further supporting that the *nfr1-2* mutation is dominant to *symRK-3* and determines an earlier step in the symbiotic signalling pathway.

25 **3. *NFR1* and *NFR5* mediated Nod-factor perception lies upstream of *NIN* and *ENOD* and is required for their expression.**

The symbiotic expression of the nodulin genes, *Lotus japonicus ENOD2* (Niwa, S. et al., 2001 *MPMI* 14:848-56) and *NIN*, in roots following rhizobial inoculation, provides a marker for *NFR* gene expression. The steady-state levels of *NIN* and *ENOD2* mRNA were measured in roots before and

following rhizobial inoculation by quantitative real-time PCR, using the primer pairs:

5' AATGCTCTTGATCAGGCTG3' (SEQ ID No: 26) and
5' AGGAGCCCAAGTGAGTGCTA3' (SEQ ID No: 27) for amplification of *N/N*
5 mRNA reverse transcripts; and the primer pairs:
5' CAG GAA AAA CCA CCA CCT GT3' (SEQ ID No: 28) and
5' ATGGAGGCGAATACACTGGTG3' (SEQ ID No: 29) for amplification of
ENOD2 mRNA reverse transcripts. The identity of the amplified sequences
was confirmed by sequencing.

10 Five hours after inoculation, induction of *N/N* gene expression was detected
in the wild type plants, while induction of *ENOD2* occurs after 12 days as
shown in Figure 10a and b. In the *nfr1* and *nfr5* mutants, activation of *N/N*
and *ENOD2* was not detected, demonstrating that functional *NFR1* and
NFR5 genes can be monitored by the activation of these early nodulin genes.

15 *Lotus* plants transformed with a *N/N* gene promoter region fused to a GUS
reporter gene provide a further tool to monitor *NFR* gene function.
Expression of the *N/N*-GUS reporter can be induced in root hairs and
epidermal cells of the root invasion zone following rhizobial inoculation in
transformed wild-type plants. In contrast expression of the *N/N*-GUS reporter
20 in an *nfr1* mutant was not detected following rhizobial inoculation. Likewise,
N/N-GUS expression was induced in the invasion zone of wildtype plants
after Nod-factor application, while in a *nfr1* mutant background no expression
was detected. The requirement for *NFR1* function was confirmed in *nfr1-1, nin*
double mutants by the absence of root hair curling and excessive root hair
25 curling (Fig 8).

The *LjCBP1* gene, T-DNA tagged with a promoter-less GUS in the T90 line,
is rapidly activated after *M. loti* inoculation as seen for *N/N*-GUS, thus
providing an independent and sensitive reporter of early nodulin gene
expression (Webb et al, 2000, Molecular Plant-Microbe Interact. 13,606,-
30 616). Parallel experiments comparing expression of the *LjCBP1* promoter
GUS fusion in wt and *nfr1* mutant background confirm the requirement for a

functional *NFR1* for activation of the early response to bacteria and Nod-factor.

Example 3.

5 Transgenic expression of NFR polypeptides and complementation of the *nfr* mutants

The *NFR* genes, encoding the *NFR1* and *NFR5* protein components of the Nod-factor binding element, can each be stably integrated, as a transgene, into the genome of a plant, such as a non-nodulating plant or a mutant non-nodulating plant, by transformation. Expression of this transgene, directed by an operably linked promoter, can be detected by expression of the respective *NFR* protein in the transformed plant and functional complementation of a non-nodulating mutant plant.

A wildtype *NFR5* transgene expression cassette of 3.5 kb, comprising a 1175 bp promotor region, the *NFR5* gene and a 441 bp 3' UTR was cloned in a vector (pIV10), and the vector was recombined into the T-DNA of *Agrobacterium rhizogenes* strain AR12 and AR1193 by triparental mating.

The *NFR5* expression cassette in pIV10 was subsequently transformed into non-nodulating *Lotus nfr5-1* and *nfr5-2* mutants via *Agrobacterium rhizogenes*-mediated transformation according to the standard protocol (Stougaard 1995, Methods in Molecular Biology volume 49, Plant Gene Transfer and Expression Protocols, p 49-63) In parallel, control transgenic *Lotus nfr5-1* and *nfr5-2* mutants plants were generated, which were transformed with an empty vector, lacking the *NFR5* expression cassette.

The nodulation phenotype of the transgenic hairy root tissue of the transformed mutant *Lotus* plants was scored after inoculation with *Mesorhizobium loti* (*M. loti*) strain NZP2235. *In planta* complementation of the *nfr5-1* and *nfr5-2* mutants by the *NFR5* transgene was accomplished, as shown in Table 2, with an efficiency of $\geq 58\%$, and the establishment of normal rhizobial-legume interactions and development of nitrogen fixing

nodules. Complementation was dependent on transformation with a vector comprising the *NFR5* expression cassette.

A transgene expression cassette, comprising the wild type *NFR1* gene comprising 3020 bp of promoter region, the *NFR1* ORF and 394 bp of 5 3' untranslated region, was cloned into the pIV10 vector and recombined into *Agrobacterium rhizogenes* strain AR12 and AR1193 by triparental mating. *Agrobacterium rhizogenes*-mediated transformation was used to transform the gene into non-nodulating *Lotus nfr1-1* and *nfr1-2* mutants in parallel with a control empty vector. *In planta* complementation of the *Lotus nfr1-1* and 10 *nfr1-2* mutants by the *NFR1* transgene was accomplished, as shown in Table 3, with an efficiency of $\geq 60\%$, and the establishment of normal *Rhizobium*-legume interactions with *M. loti* strain NZP2235, and development of nitrogen fixing nodules. Complementation was dependent on transformation with a vector comprising the *NFR1* expression cassette

15 **Example 4**

Expression and characterisation of the NFR1, NFR5 and SYM10 proteins in transgenic plants

NFR1, NFR5 and SYM10 proteins are expressed and purified from transgenic plants, by exploiting easy and well described transformation 20 procedures for *Lotus* (Stougaard 1995, supra) and tobacco (Draper et al.1988, Plant Genetic Transformation and Gene Expression, A Laboratory Manual, Blackwell Scientific Publications). Expression in plants is particularly advantageous, since it facilitates the correct folding of these transmembrane proteins and provides for correct post-translational modification, such as 25 phosphorylation. The primary sequences of the expressed proteins are extended with commercially available epitope tags (Myc or FLAG), to allow their purification from plant protein extracts. DNA sequences encoding the tags are ligated into the expression cassette for each protein, in frame, either

at the 5' or the 3' end of the cDNA coding region. These modified coding regions are then operably linked to a promoter, and recombined into *Agrobacterium rhizogenes*. *Lotus* is transformed by wound-site infection and from the transgenic roots independent root cultures are established *in vitro*

5 (Stougaard 1995, *supra*). NFR1, NFR5 and SYM10 proteins are then purified from root cultures by affinity chromatography using the epitope specific antibody and standard procedures. Alternatively the proteins are immunoprecipitated from crude extracts or from semi-purified preparations. Proteins are detected by Western blotting methods. For transformation and

10 expression in tobacco, the epitope tagged cDNAs are cloned into an expression cassette comprising a constitutively expressed 35S promoter and a 3'UTR and subsequently inserted into binary vectors. After transfer of the binary vector into *Agrobacterium tumefaciens*, transgenic tobacco plants are obtained by the transformation regeneration procedure (Draper et al. 1988, 15 *supra*). Proteins are then extracted from crude or semi-purified extracts of tobacco leaves using affinity purification or immunoprecipitation methods. The epitope tagged purified protein preparations are used to raise monospecific antibodies towards the NFR1, NFR5 and SYM10 proteins

20 **Example 5**

Plant breeding tools to select for enhanced nodulation frequency and efficiency.

A successful and efficient primary interaction between a rhizobial strain and its host depends on detection of a *Rhizobium* strain's unique Nod-factor

25 (LCO) profile by the plant host. The Nod-factor binding element and its component NFR proteins, each with their extracellular LysM motifs, play a key role in controlling this interaction. *NFR* alleles, encoding variant NFR proteins are shown to be correlated with the efficiency and frequency of

nodulation with a given rhizobial strain. Molecular breeding tools to detect and distinguish different plant *NFR* alleles, and assays to assess the nodulation efficiency and frequency of each allele, provides an effective method to breed for nodulation efficiency and frequency.

- 5 Methods useful for breeding for nodulation efficiency and frequency are given below, and the application of these techniques is illustrated for the *NFR* alleles of *Lotus* spp. Using the *Rhizobium leguminosarum* bv *viceae* 5560 DZL strain (Bras et al, 2000, Molecular Plant-Microbe Interact. 13, 475-479) it is documented that the host range of this strain within the *Lotus* spp depends
- 10 on the *NFR1* and *NFR5* alleles present in the *Lotus* host. When inoculated onto wild type plants *Rhizobium leguminosarum* bv *viceae* 5560 DZL form root nodules on *Lotus japonicus* GIFU but the strain is unable to form root nodules on *Lotus filicaulis*. Transgenic *L. filicaulis* transformed with the *Lotus japonicus* GIFU *NFR1* and *NFR5* alleles do however form root nodules when
- 15 inoculated with the *Rhizobium leguminosarum* bv *viceae* 5560 DZL strain proving the *NFR1/NFR5* allele dependent Nod-factor recognition.

1. Determining the Nod-factor specificity and sensitivity of *NFR* alleles.

Root hair curling and root hair deformation in the susceptible invasion zone is a sensitive *in vivo* assay for monitoring the legume plants ability to recognise a *Rhizobium* strain or the Nod-factor synthesized by a *Rhizobium* strain. The assay is performed on seedlings and established as follows. Seeds of wild type, transgenic and mutant *Lotus* spp are sterilised and germinated for 3 days. Seedlings are grown on 1/4 B&D medium (Handberg and Stougaard, 1992 *supra*), between two layers of sterile wet filter paper for 3 days more.

- 20 Afterwards, they are transferred into smaller petri dishes containing 1/4 B&D medium supplemented with 12.7nM AVG [(S)-trans-2-amino-4-(2-aminoethoxy)-3-butenoic acid hydrochloride] (Bras C. et al, 2000 , MPMI 13: 475-479). On transfer, the seedlings are inoculated with either 20 µl of 1:100
- 25

dilution of a 2 days old *M. loti* strain NZP2235 culture, or with *M. loti* strain R7A Nod-factor coated sand, or with sterile water as a control, and a layer of wet dialysis membrane is used to cover the whole root. A minimum of 30 seedlings are microscopically analysed for specific deformations of the root

5 hairs. The assay determines the threshold sensitivity of each *L. japonicus*, for the Nod-factor (LCO) of a given *Rhizobium* strain and the frequency of root hair curling and/or deformation.

In an alternative procedure, seeds of *Lotus japonicus* are surface sterilised and germinated for 4 days on 1% agar plates containing half-strength

10 nitrogen-free medium (Imaizumi-Anraku *et al.*, 1997, *Plant Cell Physiol.* 38: 871-881), at 26°C, under a 16h light and 8h dark regime. Straight roots, of <1cm in length, on germlings from each cultivar are then selected and transplanted on Fåhraeus slides, in a nitrogen-free medium and grown for a further 2 days. LCOs, prepared by *n*-butanol extraction and HPLC separation

15 from a given *Rhizobium* strain (Niwa *et al.* 2001, *MPMI* 14: 848-856), are applied to the straight roots in each cultivar, at a final concentration range of between 10^{-7} and 10^{-9} M. After 12 to 24h culture, the roots are stained with 0.1% toluene blue and the number of root hairs showing curling is counted. The assay determines the threshold sensitivity of each *Lotus* spp., carrying a

20 given *NFR* allele, for the Nod-factor (LCO) of a given *Rhizobium* strain and the frequency of root hair curling.

2. Determining the frequency and efficiency of nodulation of *NFR* alleles.

The efficiency of a legume plants ability to form root nodules after inoculation

25 with a *Rhizobium* strain is determined in small scale controlled nodulation tests. *Lotus* seeds are surface sterilised in 2 % hyperchlorite and cultivated under aseptic conditions in nitrogen free 1/4 concentrated B&D medium. After 3 days of germination, seedlings are inoculated with a 2 days old

culture of *M. loti* NZP2235 or TONO or R7A or with the *R. leguminosarum* bv *viceae* 5560DZL strain. In principle a set of plants is only inoculated with one stain. For controlled competition experiments where legume-*Rhizobium* recognition is determined in a mixed *Rhizobium* population, a set of plants 5 can be inoculated with more than one *Rhizobium* strain or with an extract from a particular soil. Two growth regimes are used: either petri dishes with solidified agar or Magenta jars with a solid support of burnt clay and vermiculite. The number of root nodules developed after a chosen time period is then counted, and the weight of the nodules developed can be 10 determined. The efficiency of the root nodules in terms of nitrogen fixation can be determined in several ways, for example as the weight of the plants or directly as the amount of N15 nitrogen incorporated in the plant molecules.

In an alternative procedure, *Lotus* seeds are surface sterilised and vernalised at 4°C for 2 days on agar plates and germinated overnight at 28°C. The 15 seedlings are inoculated with *Mesorhizobium loti* strain NZP2235, TONO or R7A LCOs (as described above) and grown in petri dishes on Jensen agar medium at 20°C in 8h dark, 16h light regime. The number of nodules present on the plant roots of each cultivar is determined at 3 days intervals over a period of 25 days, providing a measure of the rate of nodulation and the 20 abundance of nodules per plant.

3. Determining nodule occupancy in relation to NFR allele

In agriculture the NFR Nod-factor binding element recognises *Rhizobium* bacteria under adverse soil conditions. The final measure of a particular strain's or commercial *Rhizobium* inoculum's ability to compete with the 25 endogenous *Rhizobium* soil population for invasion of a legume crop with particular NFR alleles, is root nodule occupancy. The proportion of nodules formed after invasion by a particular strain and the fraction of the particular *Rhizobium* strain inside individual root nodules is determined by surface

sterilising the root nodule surface in hyperchlorite, followed by crushing of the nodule into a crude extract and counting the colony forming *Rhizobium* units after dilution of the extract and plating on medium allowing *Rhizobium* growth (Vincent., JM. 1970, A manual for the practical study of root nodule bacteria.

5 IBP handbook no. 15 Oxford Blackwell Scientific Publications, López-García *et al*, 2001, *J Bacteriol*, 183,7241-7252).

4. *NFR1* and *NFR5* are determinants of host range in *Lotus-Rhizobium* interactions.

10 Wild type *Lotus japonicus* Gifu is nodulated by both *Rhizobium leguminosarum* bv. *viciae* 5560 DZL (*R. leg* 5560DZL) and *Mesorhizobium loti* NZP2235 (*M.loti* NZP2235), while wild type *Lotus filicaulis* is only nodulated by *M.loti* NZP2235. Transgenic *Lotus filicaulis* plants expressing the *NFR1* and *NFR5* alleles of *Lotus japonicus* Gifu, are nodulated by *R. leg* 5560DZL, clearly demonstrating that the *NFR* alleles are the primary determinants of host range.

15

Lotus filicaulis was transformed with vectors comprising *NFR1* and *NFR5* wild type genes and their cognate promoters from *Lotus japonicus* Gifu or with empty vectors. The *Lotus filicaulis* transformants carrying *NFR1* and *NFR5* are nodulated by *R. leg* 5560DZL, albeit at reduced efficiency/frequency (9.6%) compared to *Lotus japonicus* Gifu (100%), as shown in Table 4. Mixing of NFR subunits from *Lotus japonicus* and *Lotus filicaulis* in the Nod-factor binding element is likely to contribute to the reduced efficiency observed. These data demonstrate that rhizobial strain recognition specificity is determined by the *NFR1* and *NFR5* alleles and that breeding for specific *NFR* alleles present in the germplasm or in wild relatives can be used to select optimal legume-*Rhizobium* partners.

20

25

More detailed investigations show that the rhizobial strain recognition specificity of the *NFR5* and *NFR1* alleles is determined by the extracellular domain of the *NFR5* and *NFR1* proteins. Mutant *Lotus japonicus* *nfr5* was transformed with a wild type hybrid *NFR5* gene "FinG5", encoding the

5 extracellular domain from *L. filicaulis* *NFR5* fused to the kinase domain from *L. japonicus* Gifu *NFR5* (Figure 12). The hybrid gene was operably linked to the wild type *NFR5* promoter. Control transformants, comprising wild type *L. japonicus* Gifu, *L. filicaulis* and the *Lotus japonicus* *nfr5* mutant, transformed with an empty vector, are generated in parallel. The transformed plants are

10 infected either with *M. loti* NZP2235 or with *R. leg* 5560 DZL and the formation of nodules monitored, as shown in Table 5. The FinG5 hybrid gene complements the *nfr5* mutation, and 88% of the transformants are nodulated by *M. loti* NZP2235 showing that the hybrid gene is functionally expressed. However, the *nfr5* mutants expressing the FinG5 hybrid gene are very poorly

15 nodulated by *R. leg* 5560 DZL, only 3 %, (corresponding to one plant) even after prolonged infection (40 days). This demonstrates that strain specificity of the Nod-factor binding element is determined by the extracellular domain of its component *NFR* proteins.

20 In parallel, the *Lotus japonicus* *nfr1* mutant was transformed with a wild type hybrid *NFR1* gene "FinG1", encoding the extracellular domain from *L. filicaulis* *NFR1* fused to the kinase domain from *L. japonicus* Gifu *NFR1* (Figure 12). The hybrid gene was operably linked to the wild type *NFR1* promoter. The transformed plant were infected either with *M. loti* NZP2235 or

25 with *R. leg* 5560 DZL and the formation of nodules was monitored, as shown in Table 6.

The FinG1 hybrid gene complements the *nfr1-1* mutation, and 100 % of the transformants were nodulated by *M. loti* NZP2235. However *nfr1-1* mutants expressing the FinG1 hybrid gene were less efficiently nodulated (30-40%)

30 by *R. leg* 5560 DZL. Furthermore, their nodulation by *R. leg* 5560 DZL was much delayed compared to their nodulation by *M. loti* NZP2235. Thus the

*Lotus / R. leg 5560 DZL interaction is less efficient and delayed when the transgenic host plant expresses a hybrid NFR1 comprising the extracellular domain of *Lotus filicaulis* NFR1 with the kinase domain of *Lotus japonicus* Gifu NFR1. These data indicate that the specific recognition of *R.leg 5560* DZL by its *Lotus* host is at least partly specified by the extracellular domain of NFR1 (Gifu) and that this is an allele specific recognition. However, the *NFR5* allele appears to be more important for specific recognition than *NFR1*.*

5. *NFR5* alleles and their molecular markers

10 The *NFR5* Nod-factor binding proteins encoded by the *NFR5* alleles of *Lotus japonicus* ecotype GIFU (gene sequence: SEQ ID No: 7; protein sequence: SEQ ID No: 24 & 25), and *Lotus filicaulis* (gene sequence SEQ ID No: 30; protein sequence SEQ ID No: 31) have been compared, and found to show diversity in their primary structure. Using the sequence information available

15 for the *Lotus NFR5* gene together with the pea SYM10 gene (Table 8), the alleles from different ecotypes or varieties of *Lotus*, pea and other legumes can now be identified, and used directly in breeding programs. Molecular markers based on DNA polymorphism are used to detect the alleles in breeding populations. Similar use can be taken of the *NFR1* sequences.

20 Molecular DNA markers, based on the *NFR5* allele sequence differences of *Lotus* and pea, are highlighted in Tables 8 and 9 as examples of how DNA polymorphism can be used directly to detect the presence of an advantageous allele in a breeding population.

25 Breeding for an advantageous allele can also be carried out using molecular markers, that are genetically linked to the allele of interest, but located outside the gene-allele itself. Breeding of new *Lotus japonicus* lines containing a desired *NFR5* allele can, for example, be facilitated by the use of DNA polymorphisms, (simple sequence repeats (microsatellites) or single nucleotide polymorphism (SNP) which are found at loci, genetically linked to

NFR5. Microsatellites and SNPs at the *NFR5* locus are identified by transferring markers from the general map, by identification of AFLP markers, or, by scanning the nucleotide sequence of the BAC and TAC clones spanning the *NFR5* locus, for DNA polymorphic sequences located in close proximity of the *NFR5* gene. Table 7 lists the markers closely linked to *NFR5* and the sequence differences used to design the microsatellite or SNP markers. This principle of marker assisted breeding, using genetically linked markers, can be applied to all plants. Microsatellite markers which generate PCR products with a high degree of polymorphism, are particularly useful for distinguishing closely related individuals, and hence to distinguish different *NFR5* or *NFR1* alleles in a breeding program.

Table 1**Summary of *Lotus nfr5* and pea *sym10* mutant alleles**

Allele	Mutation	<i>Lotus</i> Spp
sym5-1	EYAENGSLA 380-388 deletion	Lj
sym5-2	retrotransposon integration after Q233	Lj
sym5-3	CAG→TAG, Q55→stop	Lj
RisFixG	TGG→TGA, W ₃₈₈ →stop	Ps
P5	TGG→TGA, W ₄₀₅ →stop	Ps
P56	CAA→TAA, Q ₂₀₀ →stop	Ps
N15	<i>Sym10</i> gene deleted	Ps

TABLE 2**Complementation of *Lotus japonicus* *nfr5* mutants with the wildtype *NFR5* transgene**

<i>Lotus</i> genotype	Transgene	No. of plants	Infected With	No. of plants with nodules*	Total No. of nodules
<i>nfr5-1</i>	<i>NFR5</i>	31	<i>M. loti</i> NZP2235	18	nd
<i>nfr5-1</i>	Empty vector	20	<i>M. loti</i> NZP2235	0	nd
<i>nfr5-2</i>	<i>NFR5</i>	5	<i>M. loti</i> NZP2235	1	nd
<i>nfr5-2</i>	Empty vector	5	<i>M. loti</i> NZP2235	0	nd

* Nodules only detected on transformed roots

TABLE 3**Transformation of *Lotus japonicus* *nfr1* mutants with the wildtype *NFR1* transgene**

<i>Lotus</i> genotype	Transgene	No. of plants	Infected With	No. plants with nodules	Total No. of nodules	Average No. nodules/ plant
<i>nfr1-1</i>	<i>NFR1</i>	103	<i>M. loti</i> NZP2235	62*	310	5
<i>nfr1-1</i>	Empty vector	30	<i>M. loti</i> NZP2235	0	0	0
<i>nfr1-2</i>	<i>NFR1</i>	20	<i>M. loti</i> NZP2235	13*	97	7.5
<i>nfr1-2</i>	empty vector	7	<i>M. loti</i> NZP2235	0	0	0

* Nodules only detected on transformed roots

Table 4

***Lotus filicaulis* transformed with wildtype *NFR1* and *NFR5* genes from
Lotus japonicus Gifu**

<i>Lotus</i> genotype	Transgene	No. of plants	Infected with	No. plants with nodules	Total No. of nodules	Average No. nodules/ plant
<i>Lotus</i> <i>filicaulis</i>	<i>NFR1+</i> <i>NFR5</i>	104	<i>R.leg</i> 5560 DZL	10*	25	2.5
<i>Lotus</i> <i>filicaulis</i>	Empty vector	65	<i>R.leg</i> 5560 DZL	0	0	0
<i>Lotus</i> <i>japonicus</i> Gifu	Empty vector	10	<i>R.leg</i> 5560 DZL	10**	>150	>15

* Nodules only detected on transformed roots

** Nodules on normal and transformed roots

Table 5

L. japonicus nfr5 mutant transformed with a hybrid NFR5 gene "FinG5" encoding the extracellular domain of *L. filicaulis* NFR5 fused to the kinase domain from *L. japonicus* Gifu NFR5.

<i>Lotus</i> genotype	Transgene	No. of plants	Infected with	No. of plants with nodules	Total No. of nodules	Average No. nodules/ plant
<i>nfr5</i>	FinG5	31	<i>M. loti</i> NZP2235	28*	~180	6.4
<i>nfr5</i>	Empty vector	12	<i>M. loti</i> NZP2235	0	0	
<i>nfr5</i>	FinG5	34	<i>R. leg</i> 5560 DZL	1*	4	4 <u>1 PLANT ONLY</u>
<i>nfr5</i>	empty vector	10	<i>R. leg</i> 5560 DZL	0	0	
<i>Lotus</i> <i>japonicus</i> Gifu	empty vector	10	<i>R. leg</i> 5560 DZL	10**	>150	>15
<i>Lotus</i> <i>filicaulis</i>	empty vector	29	<i>R. leg</i> 5560 DZL	0	0	

* Nodules only detected on transformed roots

** Nodules on normal and transformed roots

Table 6

***L. japonicus nfr1* mutant transformed with a hybrid *NFR1* gene "FinG1" encoding the extracellular domain of *L. filicaulis* *NFR1* fused to the kinase domain from *L. japonicus* Gifu *NFR1*.**

<i>Lotus</i> genotype	Transgene	No. of plants	Infected with	No. of plants with nodules	Total No. of nodules	Average No. nodules/ plant
<i>nfr1-1</i>	FinG1	8	<i>M. loti</i> NZP2235	8*	59	7.3
<i>nfr1-1</i>	Empty vector	6	<i>M. loti</i> NZP2235	0	0	0
<i>nfr1-1</i>	FinG1	13	<i>R. leg</i> 5560DZL	5*#	15	3
<i>nfr1-1</i>	Empty vector	9	<i>R. leg</i> 5560DZL	0	0	0
<i>nfr1-2</i>	FinG1	10	<i>R. leg</i> 5560DZL	3*#	12	4
<i>nfr1-2</i>	Empty vector	4	<i>R. leg</i> 5560DZL	0	0	0

* Nodules only detected on transformed roots

Nodules were first counted after 56 days, while *M. loti* NZP2235 nodules were detectable after ~25 days.

Table 7**Molecular markers for *NFR5* allele breeding in *Lotus***

Marker	Genetic distance from <i>NFR5</i> locus	Lotus Ecotype	Microsatellite sequence
TM0272	2,9cM	MG-20	18xCT
		Gifu	12xCT
TM0257	1,0cM	MG-20	10xAAG
		Gifu	7xAAG
LjT13i23Sfi		Gifu	TTTGCTGCAGCAAGTCAGACTGTTAGAGGA
		Fili	TTTGCTGCAACAAGTCGGACTGTTAGAGGA
TM0522	0cM	MG-20	24xAT
		Gifu	14xAT
NFR5			
E32M54-12F	0,5cM	MG-20	TTGGAAGTTCTTTTATTAGGTTAATTAA
		Fili	TTGGAAGTTCTTTA---GGTTAATTAA
LjT01c03 Not	0,7cM	Fili	CATTCCAGAAGAAAATAAGATATAATTATG
		MG-20	CATTCCAGAAGAAAATAAGATATAATTATG
TM0168	2,2cM	Gifu	CATTCCAGAAG- AAATAAGATATAATTATG
		MG-20	19xAT
TM0021	3,8cM	Gifu	15xAT
		MG-20	16xCT
		Gifu	13xCT

Table 8

**Nucleotide sequence variation between the pea *SYM10* alleles
of pea cultivars Frisson and Finale***

Frisson	CTTGCATTTC
Finale	<u>TTCACAATTT</u> CACAACAATG GCTATCTTCT TTCTTCCTTC
Frisson	TAGTTCTCAT GCCCTTTTC TTGCACTCAT GTTTTTGTC ACTAATATTT
Finale	<u>TAGTTCTCAT</u> GCCCTTTTC TTGCACTCAT GTTTTTGTC ACTAATATTT
Frisson	CAGCTCAACC ATTACAAC TC AGTGGAACAA ACTTTTCATG CCCGGTGGAT
Finale	<u>CAGCTCAACC</u> ATTACAAC TC AGTGGAACAA ACTTTTCATG CCCGGTGGAT
Frisson	TCACCTCCTT CATGTGAAAC CTATGTGACA TACTTGCTC GGTCTCCAAA
Finale	<u>TCACCTCCTT</u> CATGTGAAAC CTATGTGACA TACTTGCTC GGTCTCCAAA
Frisson	CTTTTGAGC CTAACTAACA TATCAGATAT ATTTGATATG AGTCCTTTAT
Finale	<u>CTTTTGAGC</u> CTAACTAACA TATCAGATAT ATTTGATATG AGTCCTTTAT
Frisson	CCATTGCAAA AGCCAGTAAC ATAGAAGATG AGGACAAGAA GCTGGTTGAA
Finale	<u>CCATTGCAAA</u> AGCCAGTAAC ATAGAAGATG AGGACAAGAA GCTGGTTGAA
Frisson	GGCCAAGTCT TACTCATACC TGTAACTTGT GGTTGCACTA GAAATCGCTA
Finale	<u>GGCCAAGTCT</u> TACTCATACC TGTAACTTGT GGTTGCACTA GAAATCGCTA
Frisson	TTTCGCGAAT TTCACGTACA CAATCAAGCT AGGTGACAAC TATTCATAG
Finale	TTTCGCGAAT TTCACGTACA CAATCAAGCT AGGTGACAAC TATTCATAG
Frisson	TTTCAACCAC TTCATACCAG AATCTTACAA ATTATGTGGA AATGGAAAAT
Finale	<u>TTTCAACCAC</u> TTCATACCAG AATCTTACAA ATTATGTGGA AATGGAAAAT
Frisson	TTCAACCCCTA ATCTAAGTCC AAATCTATTG CCACCCAGAAA TCAAAGTTGT
Finale	<u>TTCAACCCCTA</u> ATCTAAGTCC AAATCTATTG CCACCCAGAAA TCAAAGTTGT
Frisson	TGTCCTTTA TTCTGCAAAT GCCCCTCGAA GAATCAGTTG AGCAAAGGAA
Finale	<u>TGTCCTTTA</u> TTCTGCAAAT GCCCCTCGAA GAATCAGTTG AGCAAAGGAA
Frisson	TAAAGCATCT GATTACTTAT GTGTGGCAGG CTAATGACAA TGTTACCCGT
Finale	<u>TAAAGCATCT</u> GATTACTTAT GTGTGGCAGG CTAATGACAA TGTTACCCGT
Frisson	GTAAGTTCCA AGTTGGTGC ATCACAAGTG GATATGTTA CTGAAAACAA
Finale	<u>GTAAGTTCCA</u> AGTTGGTGC ATCACAAGTG GATATGTTA CTGAAAACAA
Frisson	TCAAAACTTC ACTGCTTCAA CCAACGTTCC GATTTTGATC CCTGTGACAA
Finale	<u>TCAAAACTTC</u> ACTGCTTCAA CCAATGTTCC GATTTTGATC CCTGTGACAA

Frisson	AGTTACCGGT AATTGATCAA CCATCTCAA ATGGAAGAAA AAACAGCACT
Finale	<u>AGTTACCGGT AATTGATCAA CCATCTCAA ATGGAAGAAA AAACAGCACT</u>
Frisson	CAAAAACCTG CTTTTATAAT TGGTATTAGC CTAGGATGTG CTTTTTCGT
Finale	<u>CAAAAACCTG CTTTTATAAT TGGTATTAGC CTAGGATGTG CTTTTTCGT</u>
Frisson	TGTAGTTTA ACACTATCAC TTGTTTATGT ATATTGTCTG AAAATGAAGA
Finale	<u>TGTAGTTTA ACACTATCAC TTGTTTATGT ATATTGTCTG AAAATGAAGA</u>
Frisson	GATTGAATAG GAGTACTTCA TTGGCGGAGA CTGCGGATAA GTTACTTTCA
Finale	<u>GATTGAATAG GAGTACTTCA TTGGCGGAGA CTGCGGATAA GTTACTTTCA</u>
Frisson	GGTGTTCGG GTTATGTAAG CAAGCCAACA ATGTATGAAA TGGATGCGAT
Finale	<u>GGTGTTCGG GTTATGTAAG CAAGCCAACA ATGTATGAAA TGGATGCGAT</u>
Frisson	CATGGAAGCT ACAATGAACC TGAGTGAGAA TTGTAAGATT GGTGAATCGG
Finale	<u>CATGGAAGCT ACAATGAACC TGAGTGAGAA TTGTAAGATT GGTGAATCGG</u>
Frisson	TTTACAAGGC TAATATAGAT GGTAGAGTT TAGCAGTGAA AAAATCAAG
Finale	<u>TTTACAAGGC TAATATAGAT GGTAGAGTT TAGCAGTGAA AAAATCAAG</u>
Frisson	AAAGATGCTT CTGAGGGAGCT GAAAATTCTG CAGAAGGTAA ATCATGGAAA
Finale	<u>AAAGATGCTT CTGAGGGAGCT GAAAATTCTG CAGAAGGTAA ATCATGGAAA</u>
Frisson	TCTTGTGAAA CTTATGGGTG TGTCTTCGA CAACGACGGGA AACTGTTCC
Finale	<u>TCTTGTGAAA CTTATGGGTG TGTCTTCGA CAACGACGGGA AACTGTTCC</u>
Frisson	TTGTTTACGA GTATGCTGAA AATGGATCAC TTGATGAGTG GTTGTCTCA
Finale	<u>TTGTTTACGA GTATGCTGAA AATGGATCAC TTGATGAGTG GTTGTCTCA</u>
Frisson	GAGTGTGCGA AAACCTCGAA CTCGGTGGTC TCGCTTACAT GGTCTCAGAG
Finale	<u>GAGTGTGCGA AAACCTCGAA CTCGGTGGTC TCGCTTACAT GGTCTCAGAG</u>
Frisson	AATAACAGTA GCAGTGGATG TTGCAATAC ATGCATGAAC
Finale	<u>AATAACAGTA GCAGTGGATG TTGCAATAC ATGCATGAAC</u>
Frisson	ATACTTACCC AAGAATAATC CACAGAGACA TCACAAACAAG TAATATCCTT
Finale	<u>ATACTTACCC AAGAATAATC CACAGAGACA TCACAAACAAG TAATATCCTT</u>
Frisson	CTGGATTCAA ACTTTAAGGC CAAGATAGCG AATTTTCAA TGGCCAGAAC
Finale	<u>CTGGATTCAA ACTTTAAGGC CAAGATAGCG AATTTTCAA TGGCCAGAAC</u>
Frisson	TTCAACAAAT TCCATGATGC CGAAAATCGA TGTTTCGCT TTTGGGGTGG
Finale	<u>TTCAACAAAT TCCATGATGC CGAAAATCGA TGTTTCGCT TTTGGGGTGG</u>
Frisson	TTCTGATTGA GTTGCTTACC GGCAAGAAAG CGATAACAAAC GATGGAAAAT
Finale	<u>TTCTGATTGA GTTGCTTACC GGCAAGAAAG CGATAACAAAC GATGGAAAAT</u>

Frisson	GGCGAGGTGG TTATTCTGTG GAAGGATTTC TGGAAGATTT TTGATCTAGA
Finale	<u>GGCGAGGTGG TTATTCTGTG GAAGGATTTC TGGAAGATTT TTGATCTAGA</u>
Frisson	AGGGAATAGA GAAGAGAGCT TAAGAAAATG GATGGATCCT AAGCTAGAGA
Finale	<u>AGGGAATAGA GAAGAGAGCT TAAGAAAATG GATGGATCCT AAGCTAGAGA</u>
Frisson	ATTTTATCC TATTGATAAT GCTCTTAGTT TGGCTTCTTT GGCAGTGAAT
Finale	<u>ATTTTATCC TATTGATAAT GCTCTTAGTT TGGCTTCTTT GGCAGTGAAT</u>
Frisson	TGTACTGCAG ATAAATCATT GTCAAGACCA AGCATTGCAG AAATTGTTCT
Finale	<u>TGTACTGCAG ATAAATCATT GTCAAGACCA AGCATTGCAG AAATTGTTCT</u>
Frisson	TTGTCTTTCT CTTCTCAATC AATCATCATC TGAACCAATG TTAGAAAGAT
Finale	<u>TTGTCTTTCT CTTCTCAATC AATCATCATC TGAACCAATG TTAGAAAGAT</u>
Frisson	CCTTGACATC TGGTTAGAT GTTGAAGCTA CTCATGTTGT TACTTCTATA
Finale	<u>CCTTGACATC TGGTTAGAT GTTGAAGCTA CTCATGTTGT TACTTCTATA</u>
Frisson	GTAGCTCGTT GATATTCAATT CAAGTGAAGG TAACACTGAA TCAATGCTTC
Finale	<u>GTAGCTCGTT GATATTCAATT CAAGTGAAGG TAACACTGAA TCAATGCTTC</u>
Frisson	AGTTTCTTAT ATTCAAGATG GTTACTTTGT TTAGATGATT ATTGATTACA
Finale	<u>AGTTTCTTAT ATTCAAGATG GTTACTTTGT TTAGATGATT ATTGATTACA</u>
Frisson	TCTTTATGTG TGGAACATAA TGTTATTTT AATTAAGGGA ATTGTCTAA
Finale	<u>TCTTTATGTG TGGAACATAA TGTTATTTT AATTAAGGGA ATTGTCTAA</u>
Frisson	AATTCATTT TCCATGTT
Finale	<u>AATTCATTT TCCATGTT</u>

* Nucleotide differences are shaded black and the coding region is underlined

Table 9

**Protein sequence differences encoded by the pea *SYM10* alleles
of pea cultivars Frisson and Finale***

Frisson	MAIFFLPSSS	HALFLALMFF	VTNISAQPLQ	LSGTNFSCPV	DSPPSCETYV
Finale	MAIFFLPSSS	HALFLALMFF	VTNISAQPLQ	LSGTNFSCPV	DSPPSCETYV
Frisson	TYFARSPNFL	SLTNISDIFD	MSPLSIAKAS	NIEDEDKKLV	EGQVLLIPVT
Finale	TYFARSPNFL	SLTNISDIFD	MSPLSIAKAS	NIEDEDKKLV	EGQVLLIPVT
Frisson	CGCTRNRYFA	NFTYTIKLGD	NYFIVSTTSY	QNLTNYVEME	NFNPNLSPNL
Finale	CGCTRNRYFA	NFTYTIKLGD	NYFIVSTTSY	QNLTNYVEME	NFNPNLSPNL
Frisson	LPPEIKVVVP	LFCKCPSKNQ	LSKGIKHLIT	YVWQANDNVT	RVSSKFGASQ
Finale	LPPEIKVVVP	LFCKCPSKNQ	LSKGIKHLIT	YVWQANDNVT	RVSSKFGASQ
Frisson	VDMFTENNQN	FTASTNVPIL	IPVTKLPVID	QPSSNGRKNS	TQKPAIFIIGI
Finale	VDMFTENNQN	FTASTNVPIL	IPVTKLPVID	QPSSNGRKNS	TQKPAIFIIGI
Frisson	SLGCAFFVVV	LTLSLVYVYC	LKMKRRLNRST	SLAETADKLL	SGVSGYVSKP
Finale	SLGCAFFVVV	LTLSLVYVYC	LKMKRRLNRST	SLAETADKLL	SGVSGYVSKP
Frisson	TMYEMDAIME	ATMNLSENCK	IGESVYKANI	DGRVLAVKKI	KKDASEELKI
Finale	TMYEMDAIME	ATMNLSENCK	IGESVYKANI	DGRVLAVKKI	KKDASEELKI
Frisson	LQKVNHGGLV	KLMGVSSDND	GNCFLVYeya	ENGSLDEWL	SESSKTSNSV
Finale	LQKVNHGGLV	KLMGVSSDNE	GNCFLVYeya	ENGSLDEWL	SELSKTSNSV
Frisson	VSLTWSQRIT	VAVDVAVGLQ	YMHEHTYPRI	IHRDITTSNI	LLDSNFKAKI
Finale	VSLTWSQRIT	VAVDVAVGLQ	YMHEHTYPRI	IHRDITTSNI	LLDSNFKAKI
Frisson	ANFSMARTST	NSMMPKIDVF	AFGVVLIELL	TGKKAITTME	NGEVVILWKD
Finale	ANFSMARTST	NSMMPKIDVF	AFGVVLIELL	TGKKAITTME	NGEVVILWKD
Frisson	FWKIFDLEGN	REESLRKWM	PKLENFYPID	NALSLASLAV	NCTADKSLSR
Finale	FWKIFDLEGN	REESLRKWM	PKLENFYPID	NALSLASLAV	NCTADKSLSR
Frisson	PSIAEIVLCL	SLLNQSSSEP	MLERSLTSGL	DVEATHVVTS	IVAR
Finale	PSIAEIVLCL	SLLNQSSSEP	MLERSLTSGL	DVEATHVVTS	IVAR

* Amino acid differences are highlighted in black.

Claims

1. A Nod-factor binding element comprising one or more isolated NFR polypeptides.
- 5 2. The Nod-factor binding element of claim 1, wherein the NFR polypeptide is NFR1, comprising an amino acid sequence substantially identical to SEQ ID No: 25, and having specific Nod-factor binding properties.
- 10 3. The Nod-factor binding element of claim 1, wherein the NFR polypeptide is NFR5 comprising an amino acid sequence substantially identical to SEQ ID No: 8 and having specific Nod-factor binding properties.
- 15 4. The Nod-factor binding element of claim 1, comprising the NFR polypeptides NFR1 and NFR5, comprising amino acid sequences substantially identical to SEQ ID No: 25 and SEQ ID No: 8, respectively, and having specific Nod-factor binding properties.
5. The Nod-factor binding element of claim 2 or 4, wherein said NFR1 polypeptide is encoded by an *NFR1* gene comprising a nucleic acid sequence substantially identical to SEQ ID No: 23.
- 20 6. The Nod-factor binding element of claim 3 or 4, wherein said NFR5 polypeptide is encoded by an *NFR5* gene comprising a nucleic acid sequence substantially identical to SEQ ID No: 7.
7. An isolated polynucleotide molecule comprising a nucleic acid sequence encoding the NFR1 polypeptide of claim 2 that is substantially identical to SEQ ID No: 23.
- 25 8. An isolated polynucleotide molecule comprising a nucleic acid sequence encoding the NFR5 polypeptide of claim 3 that is substantially identical to SEQ ID No: 7.
9. A method of producing a plant expressing the Nod-factor binding element of claims 2 or 3, the method comprising introducing into the plant a transgenic expression cassette comprising a nucleic acid

sequence encoding a NFR polypeptide that is substantially identical to SEQ ID No: 25 or SEQ ID No:8, and having specific Nod-factor binding properties, wherein the nucleic acid sequence is operably linked to a promoter and selecting transgenic plants and their progeny expressing said NFR polypeptide.

5 10. The method of claim 9, wherein the transgenic expression cassette is introduced into the plant through a sexual cross.

11. The method of claim 9, wherein said promoter is a native promoter or heterologous root specific promoter.

10 12. The method of claim 9, wherein said promoter is a native or heterologous constitutive promoter.

13. A transgenic plant expressing one or more NFR polypeptides produced according to the method of any one of claims 9 to 12.

14. The transgenic plant of claim 13, expressing NFR 1 and NFR5

15 polypeptides comprising an amino acid sequence substantially identical to SEQ ID No: 25 and SEQ ID No: 8, respectively and having specific Nod-factor binding properties.

15. The transgenic plant of claim 13, wherein the plant is a non-nodulating dicotyledenous plant.

20 16. The transgenic plant of claim 13, wherein the plant is a monocotyledonous cereal.

17. A method for marker assisted breeding of *NFR* alleles, encoding variant NFR polypeptides, comprising the steps of:

25 a. identifying variant NFR1 or NFR5 polypeptides in a nodulating legume species, comprising an amino acid sequence substantially similar to SEQ ID No: 25 or SEQ ID No: 8 respectively, having specific Nod-factor binding properties, and

b. determining the nodulation frequency of legume plants expressing said variant NFR1 or NFR5 polypeptide, and

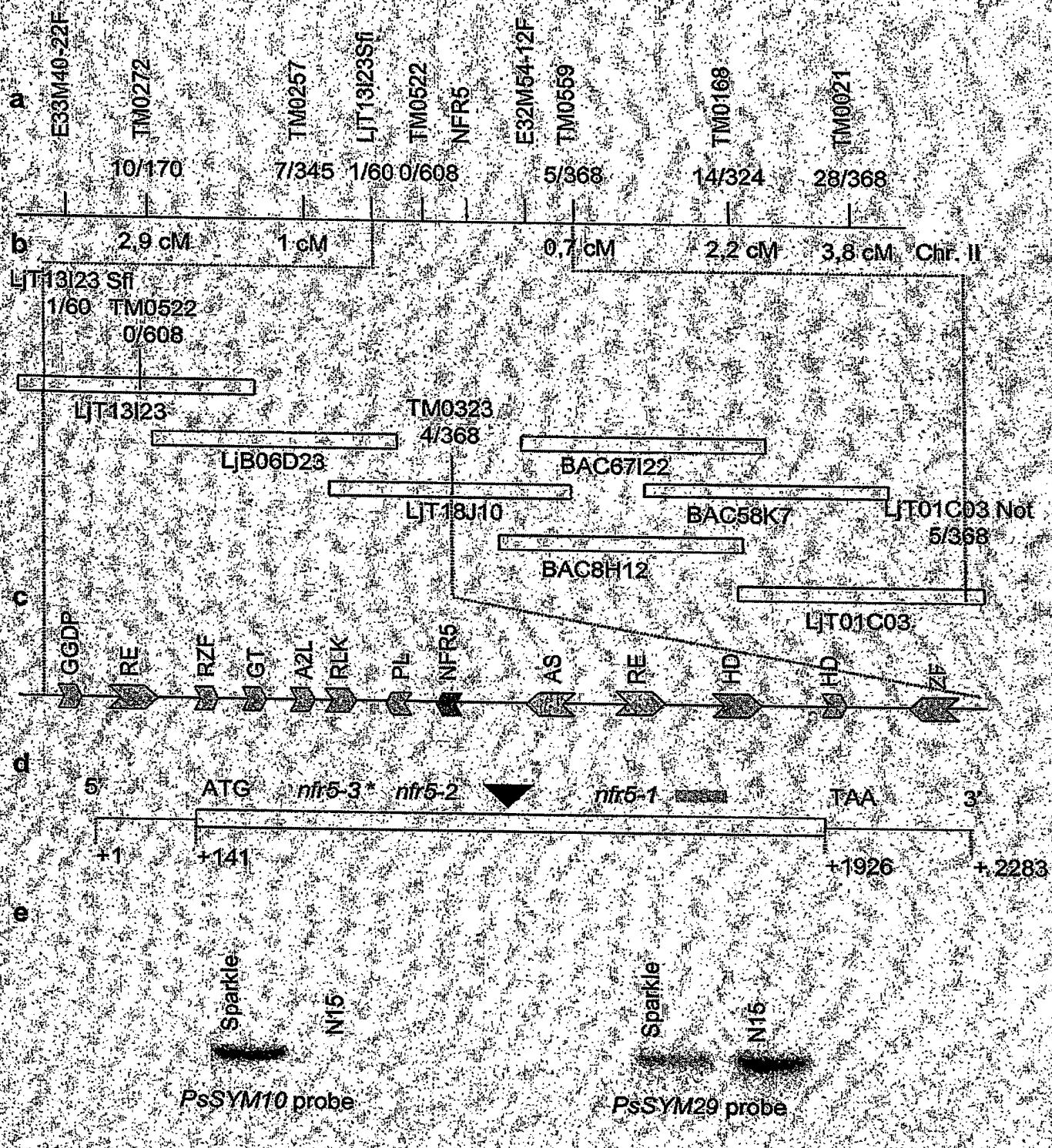
- c. identifying DNA polymorphisms at loci genetically linked to or within the allele encoding said variant NFR1 or NFR5 polypeptide, and
 - d. preparing molecular markers based on said DNA polymorphisms, and
 - e. using said molecular markers for the identification and selection of plants carrying *NFR* alleles encoding said variant NFR1 or NFR5 polypeptides.
- 5 18. Plants selected according the method of claim 17, carrying *NFR* alleles encoding variant NFR1 or NFR5 polypeptides comprising an amino acid sequence substantially similar to SEQ ID No: 25 or SEQ ID No: 8, respectively, and having specific Nod-factor binding properties.
- 10 19. Use of the method of claim 16 for breeding legumes with enhanced nodulation frequency or root nodule occupancy or enhanced symbiotic nitrogen fixation ability.
- 15

Abstract

The present invention provides a Nod-factor binding element, comprising one or more NFR polypeptides encoded by *NFR* genes, that are useful for providing non-nodulating plants with Nod-factor binding properties and 5 triggering the endosymbiotic signalling pathway leading to nodulation. Furthermore the invention is useful for breeding for improved nodulation in nodulating legumes.

03 JULI 2003

Modtaget



03 JULI 2003

Modtaget

Figure 2

a

SP LysM1 LysM2 LysM3 TM Kinase

MAVEFLTSGSLSLFIALTLIIFTNIAA SP 51 aa

RSEKISGPDFSCPVDSPSPPSCTYVT

XTAQSPNLLSLTNISDIFDISPLSIARASNIDAGKDKLIVPGOVLLVP LysM1

VTCCGCAGNHSSANTSS 113 aa

YOTOLGDSYDEVATTLYENLTNWNTIVQASNPGVNPYLLPERVKVVFPL LysM2

LFCRCPSKNOLNKGIQYLIT 180 aa

YVWKPNDNVSLVSAKFGASPADILTENRYGODFTAATNLPIIIP LysM3

VTOLPELTQPSNSGRKSSIHL 246 aa

VILGITLGCTLLTAVLTGTLVYVYC TM

RRKKALNRTASSAETADKILSGVSGYVSKPNVYEIDEI KD

I

MEATKDESDECKVGESVYKANIEGRVVAVKKIKEGGANEELKILOKV

II III

NHGNLVKLMGVSSGYDGNCFLVVYEYAENGSLAEWLFSKSSGTPNSLT

IV V VI

WSORISIAVDVAVGLOYMHEHTYPRIIHRDITTSNILEDNSNEKAKIA

VII IX

NFAMARTSLNPMMPKIDVFAFGVLLIELITGRKAMTTKENGEVVMLW

KDMWEIFDIEENREERIRKWMDPNLESFYHIDNALSLASLAVNCTAD

XI

KSLGRPSMAEIVLSLSFLTOOSSNPTLERSLTSSGLDVEDDAHTVTS

ITAR 595 aa

03 JULI 2003

Modtaget

Figure 2

c

NFR5M1	52:	YTAQOSPNLISLNIISDFFPISPLSIARASNIIDAGKDXKLIPQQLLVP	: 98
SYM10M1	52:	YFAFSPNRLISLNIISDFFDMSPLSIAKASNTEDEEKRLIPQQLLVP	: 98
M_tM1	53:	YRAQOSPNLISLNSNISDFNLSPERTAKASNTEDEEKRLIPQQLLVP	: 99
RiceM1	47:	YRTQSPGFLDLGNISDLFGVSRALIASANKLTDGVVLLRQQLLVP	: 93
NFR5M2	114:	YDQOLGDSYDEVATILYENLTNWNIYQASNPQVNPYLLPDRVWVWFP	: 160
SYM10M2	114:	YTTKLGDNYFIVSITTSYONLTNYVEMENFNPNLSENELEPEIKVWFP	: 160
M_tM2	115:	YSTKQGDNPFIFLSITTSYONLTNYLEFKNFPNLSPTHLPLDTKVSVP	: 161
RiceM2	109:	YPMRPRDTTEGELAVIAFENLTDDEVILVEETNPAAATRLEPWCEVWFP	: 155
VolvoxM2	106:	YTHQPGDTIWATAQORRGHTVVDVIQSIINPQVNPYARLOVQCOVINV	: 149
Pfam	1:	YIVVKKGDILWKTARRVGGTISVSLKSTNCGSSDNLYVCGKELKTP	: 43
NFR5M3	181:	YVWKPNNDIVSIVSAKRGASPADTITENRYGDDTTATNPLTILIP	: 224
SYM10M3	181:	YVWQANDIVTRVSSKFGASQVDMFTENNNMETASTNVPILIP	: 222
M_tM3	182:	YVWQDNDIVTIVSSKFGASQVEMTAENNNHETASTNRSVILIP	: 223
RiceM3	176:	YVWQPGDDVSIVSALMNASAANIAASNGVAGNSTATGQPVILIP	: 219

d

	VII	VIII	IX
cons	DFG	APE	D.W.G.
Smart	195: KIADEGLSP	DLYSDYYKVKGGKLPIRWMAESIKEKFTSKSDVWSFG	: 248
Arab	500: KIAANFGVARILDEGDLBLOLTRHVEGTQGYLAPEYVENGVITSKLDVFAFG	: 550	
NFR5	448: KIAANFGMARTSTN	PMMPKIDVFAFG	: 472
SYM10	449: KIAANFGSMARTSTN	SMMPKIDVFAFG	: 473
M_t	450: KIAANFGMARTSTN	SMMPKIDVFAFG	: 474
Rice	476: KLSNFSLAKPAAMV	AAATSSDVFAGF	: 502

03 JULI 2005

Modtaget

Figure 3

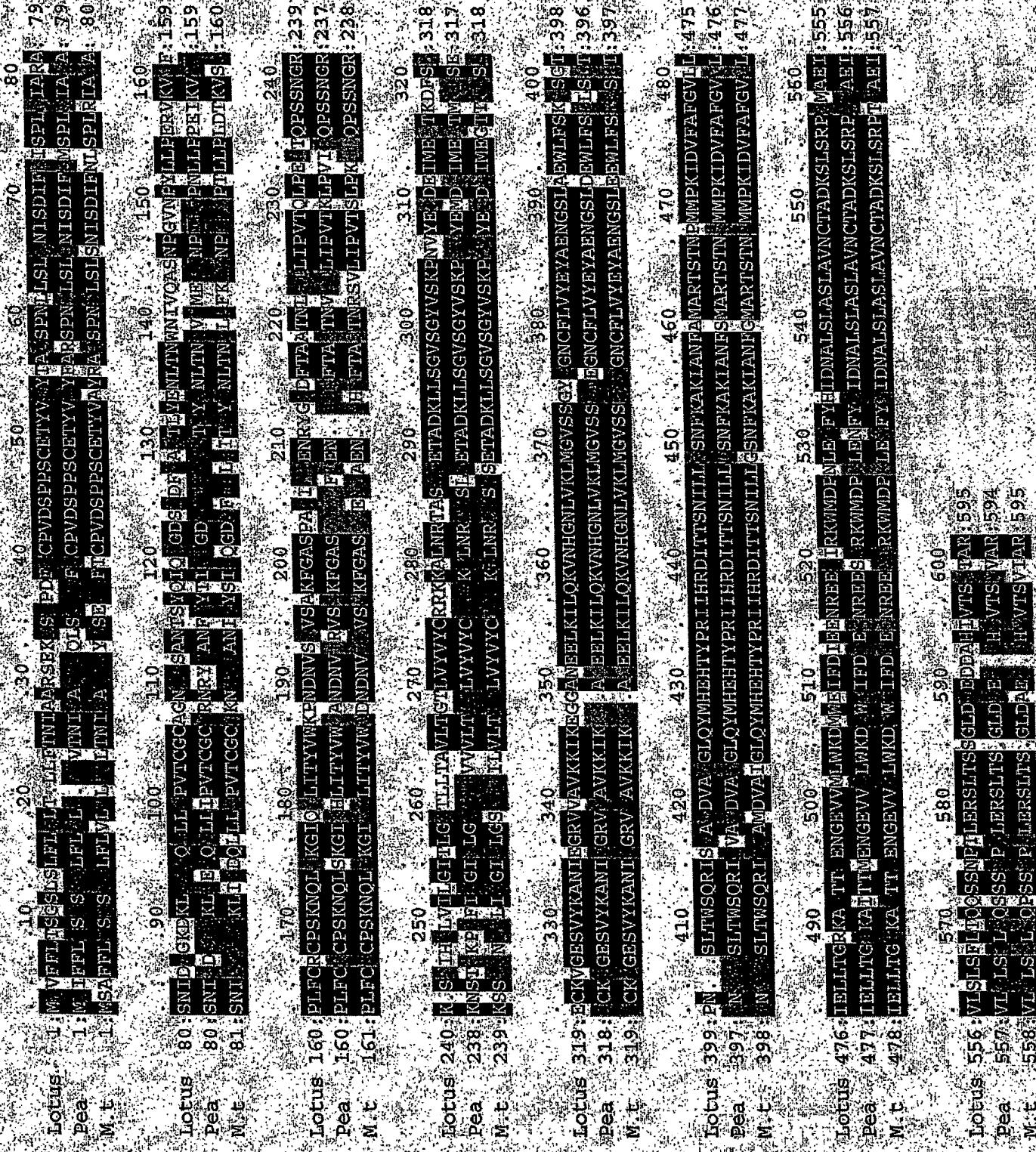
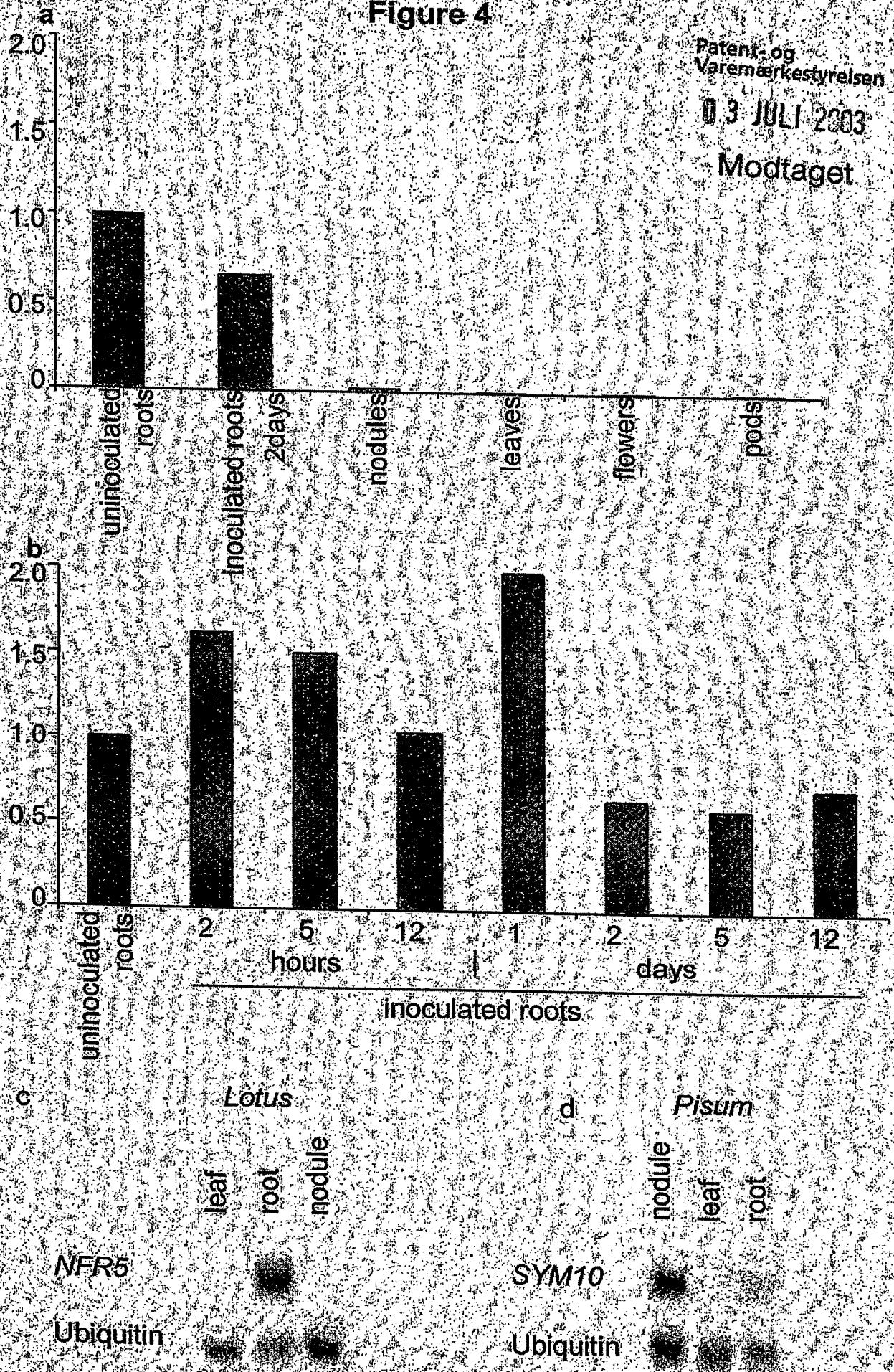


Figure 4

Patent og
Varemærkestyrelsen

03 JULI 2003

Modtaget



03 JULI 2003

Medtaget

03 JULI 2003

Modtaget

Figure 6

a

MKLKTGLLLFFILLLGH

SP

VCFHVIESNCLKGCDLALASYYILPGVFILQNTTMOSEIVSSNDAITS
YNKD**KILLNDINIOSFORLNIPFP**CDCTIGGEFLGHVFE

103

YSASKGDTIYETTIANLYANLTVDILLKRFNSYDPKNIPVNAKVNV

LysM1

VNCSCGNSOVSKDYGLFIT

168

YPIIRPGDTILODIANQSSLDAGLIQSFNPSVNFSKDSGIAFIP

LysM2

GRYKNGVYVPLYHR

224

TAGLASGAAVGISIAGTFVILLLAFCMYV

TM

RYOKKEEEKAKLPTDISMALSTODA (GN) SSSAEYETSGSSCPGTASAT
GLTSIMVAKSMEFSYOEELAKATNN

322 (324)

FSLDNKIGOGGFGAVYYAELRGKKTAIKKMDVCASTEFLCELKVLTHV

I

II

III

KD

HHLNIVRITGYCVEGSILFLVYEHDNGNLGQVLHGSGKEPLPWSSRVQTA

IV

V

VIa

LDAARGLEYTHEHTVPVYIHRDVKSANTIIDKNLRGKVADFG**LT**KLIEVG

Via

VIb

VII

NSTIOTRIVGTFGYMPPEYAOYGDISPKIDVYAFGVVLFELTSAKNAVLT

621 (623)

VIII

IX

GELVAESKGLVALFEAINKSDPCDALRKLVDPRLGENYPIDSVLIKIAOLG

*

RACTIRDNPILLRPSMRSLVVALMTISSLTEDCDESSYESOTLINLLSVR*

XI

b

SMART0257

YIKKEDTSSARRGISVS--LEINNIIIDNLOGOKKIP-

NFR1-M1

104 YIASDTATIANL--VILKRENSYDKNVNSVLT--149

At21630-M1

105 YIROEDTIVVALSN--VMESTQABNPFPATNLSLT--LV-151

SMART0257

YIVKKEDSSSIARRYIIVSILLENIN-ILDIDNLOVOKKIP

NFR1-M2

167 YPIRGTQDQDIANQSSLDAGLIQSFN-SFSKDSG-IAFP-208

At21630-M2

170 YPLRIFDSSSIARSSIVADITRYN-GFNSCNG-VVYP-211

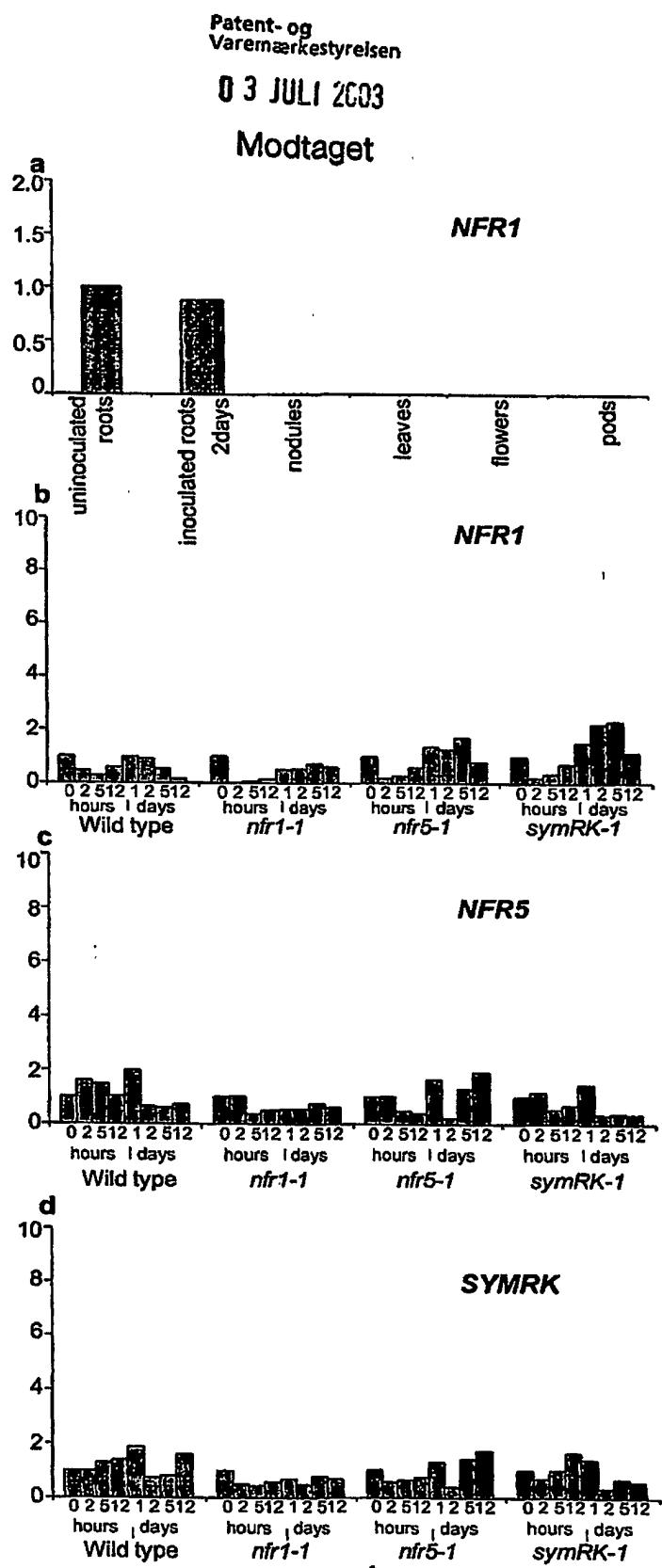
BAB89226-M2

168 YAVQEDGNTASLERSWKLTENRVAIDFIKEWILFP-212

VOLVOX M

42 YTTQEDFWAIAORRITTVDMV-SN-GMARLOVQVINVE-85

Figure 7

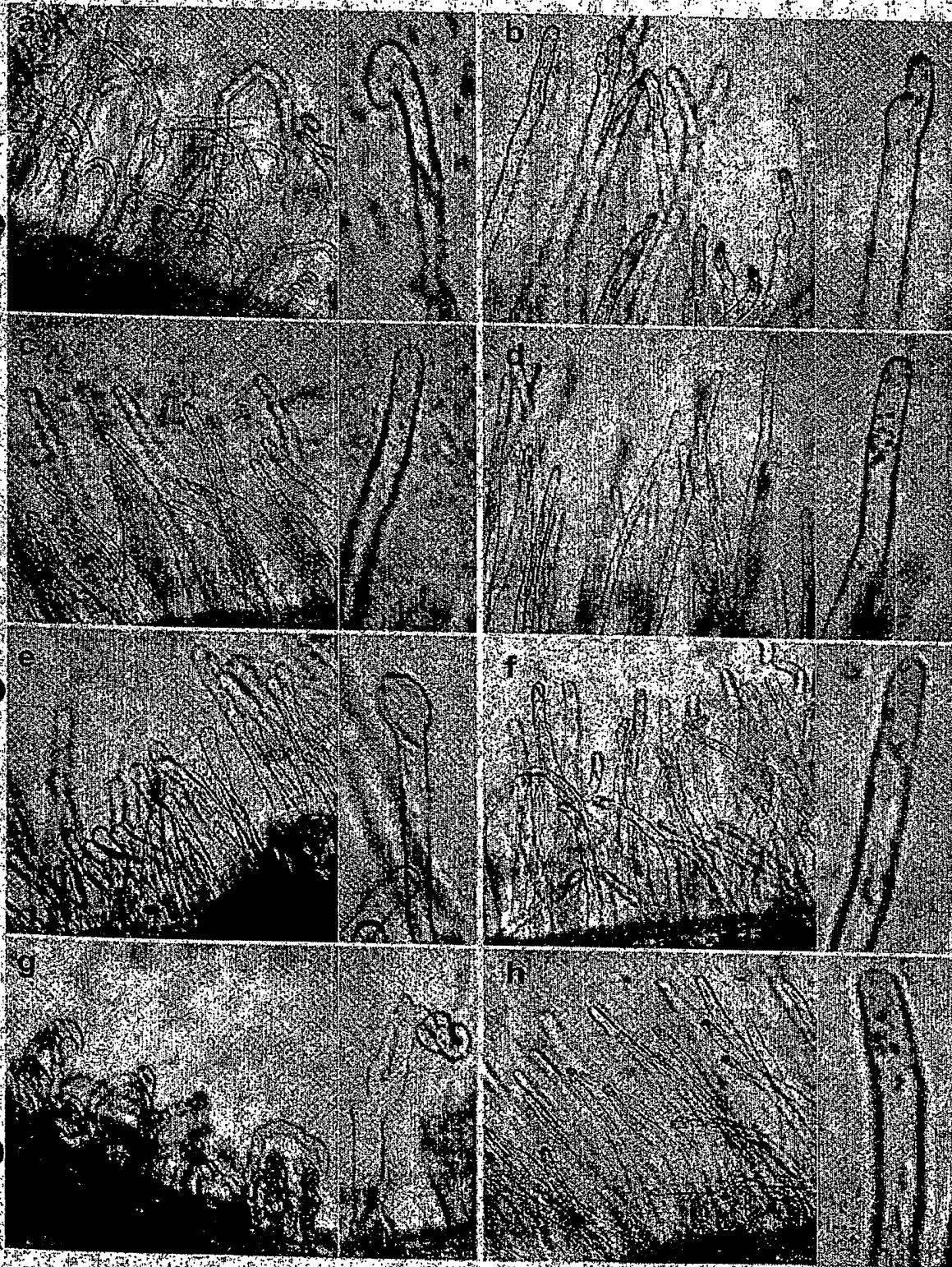


Patent- og
Varemærkestyrelsen

03. JULI 2003

Modtaget

Figure 8

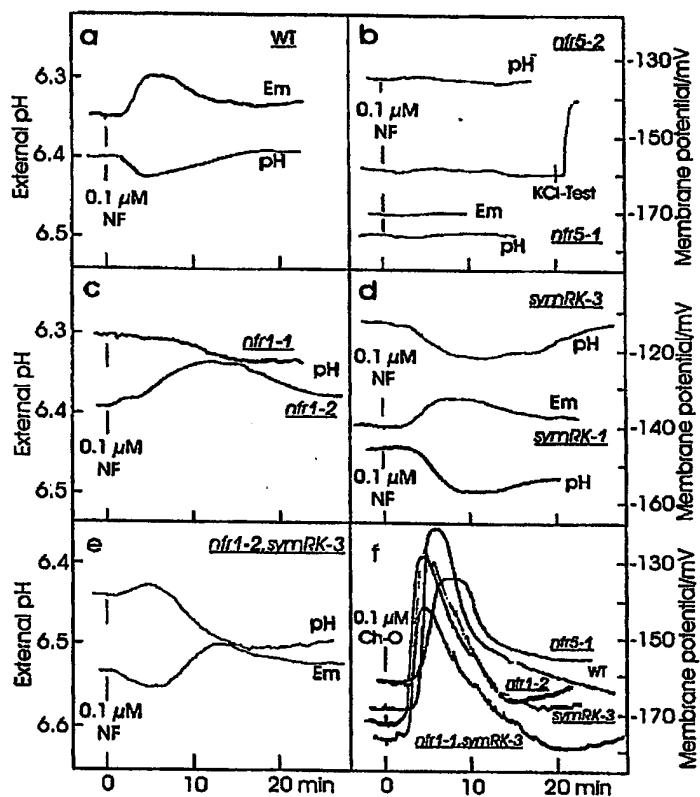


Patent- og
Varemærkestyrelsen

03 JULI 2003

Modtaget

Figure 9



Patent- og
Varemærkestyrelsen

03 JULI 2003

Modtaget

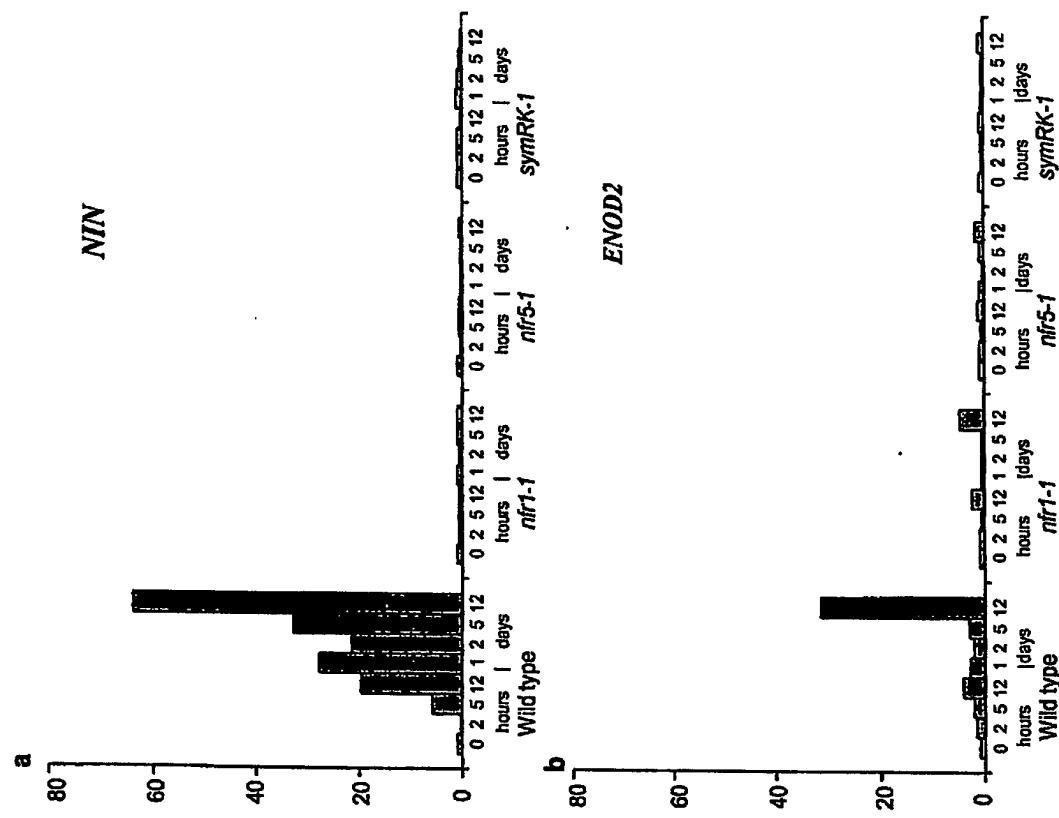


Figure 10

Figure 11

NFR1 1: MK...[KTKG.][LFFF][JCHVCFHESNCLKG]...C..D..[LALASYYL...[PGFELIONITTMQSEIIVSNDAITSYNKKDYLNDINIQSFORL
NFR5 1:[MAVFFLTSLSLFLA][LFTNIAARSEKISGPDFSPVDSPPSCETIVTYTAQSPNLLSUTNISDIEDISPLSIARAS..[NIDRGK. DKLVPG. QVL

NFR1 85: NIEPPCDICIGBEHIGHYFEY SASKEDTYETIANLYYPA[LTVDLILKRN. SYDPKNIPVNAKVNVTWNCSCGNS. QVS[KDYGFLITYPTIRPCDTLQDIAN
NFR5 96: LVPVTCGCAGNHSANT. SMIQIQLGDSYDFVATLLE[LTNWNIVQASNPGVNPYLLPERVVKVFPFLFCRCPSKNQLNKGQYLITYWKKENENVSLUSA

NFR1 183: QSSLDAGLIQSFN. PSVNESSKDSGI. AFIPGRYKNGVYVFLYHRTAGLAS[AAVGISIAGTFVLLAFCMIVRYQKKEEAKLPTDISWALSTQDASS
NFR5 195: KRGASPADIL[ENRYGOLFATANLPLIP...VT.: OLFELTOPS..SNERKS..SIHLLVIGITLGCTLL. TAVITGTLVYVYCRRKKA[IN. RTASS

NFR1 281: SAEYETSGSSCPGETASATGLITSIVAKSMEFSYQELAKATNNFSLDNKIGQGGF[GAVY]AELRGKKTAAKMDVOA[STEFLCKVLT[HHNLVPLIG
NFR5 283: AETADKLISG...VSGY...VSKPENYEDIMEAKD[ESDECKVGEES...VYKANIECRVVAVKKIKKEGGANE...ELKILQKVNHGNLWKG

NFR1 381: YC..VEGSISFLVYHLDNGN[QYQVH]G..SGKEP..LFWSSSRVQIAID[DA[RGLEYIHEHTVPIIHRDVKSANLILDKNLRGVADFGITKLIEVGNSFLQ
NFR5 367: VSSGTDGNCFLVVEYAENGSLIAEW[IFS]SSGT[PNSSL]WTSQRISIAV[DEAVGLOMMEHTYPIIHRDITTSNLLDSTPKAKIANFAMAR...TST..

NFR1 476: TRIVGTGYMPPEYAQYGDISEKIDVYAFGVVLFELISAKNVL..KIGELVAESKG[GLVALFEA[INKSDPCDALARLKVDERLGENYPIDSV[KLIAQICGR
NFR5 459:NP.....MMEKIDVYAFGVVLLIELLTGRKAMT[TKENG]VWMLWODWETDIEENR..EERIKWMDE[ENLESF]YHIDNALSLA[IAV

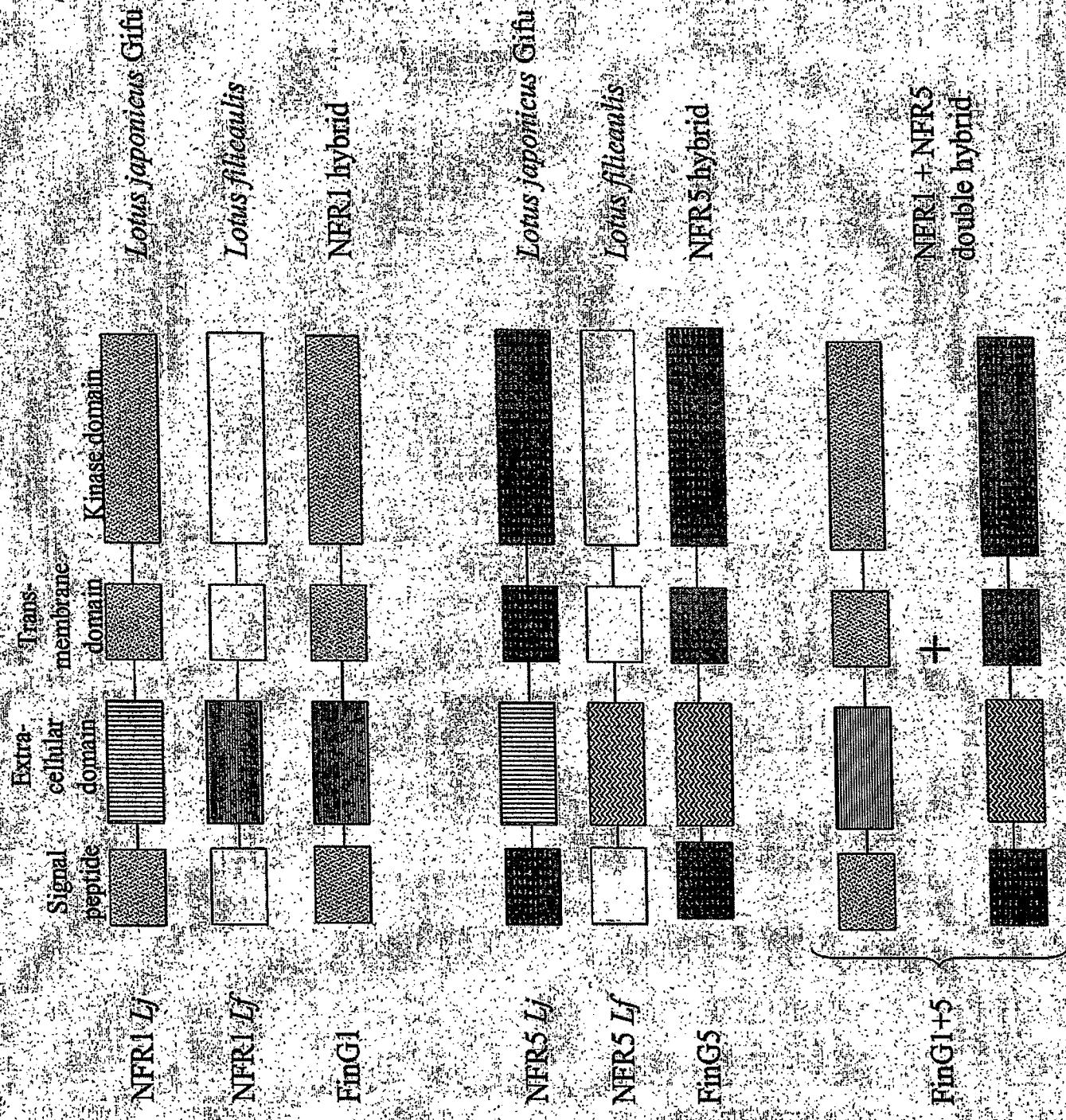
NFR1 574: ACTRDNPPLRPSMRSI[VALM[ISSLT3DDESSY[.....OTLINLISVR
NFR5 540: NCTADKSLSRPSM[AEIVL[SLFUTQOSSNPTERSLTSGLDVEDDAHIVTSITAR

Patent- og
Varemærkestyrelsen

03 JULI 2003

Modtaget

Protein domain structure of *Lotus japonicus* and *Lotus falcatus* NFRI and NFRS proteins and of the hybrid proteins



SEQUENCE LISTING

Patent- og
Varemærkestyrelsen

03 JULI 2003

Modtaget

<110> Østjysk Innovation A/S

<120> Nod-factor perception

<130> P200300646 DK

<160> 31

<170> PatentIn version 3.1

<210> 1

<211> 45

<212> DNA

<213> Lotus japonicus

<400> 1

ctaatacgcac tcactatagg gcaaggcagt gtaacaacgc agagt

45

<210> 2

<211> 29

<212> DNA

<213> Lotus japonicus

<400> 2

gctagttaaa aatgtaatag taaccacgc

29

<210> 3

<211> 21

<212> DNA

<213> Lotus japonicus

<400> 3

aaaggcagcat tcatcttctg g

21

<210> 4

<211> 39

<212> DNA

<213> Artificial

<400> 4

gaccacgcgt atcgatgtcg actttttttt ttttttttv

39

<210> 5

<211> 19

<212> DNA

<213> Lotus japonicus

<400> 5

gcaagggaaag gtaattcag

19

<210> 6

<211> 2292

<212> DNA
<213> Lotus japonicus

<400> 6
ttattgatat actaaaccac aggatattt attgacaatg tgaatgtcc atatttcaa 60
caatgctgat tccctctgat aaagaacaag ttcctttct cttccctgt taactatcat 120
ttgtccccca cttcacaaac atggctgtct tctttttac ctctggctct ctgagtc 180
ttcttgcact cacgttgctt ttcactaaca tcgcccgtcg atcagaaaaag attagcggcc 240
cagacttttc atgcctgttt gactcacctc cttcttgta aacatatgtg acatacacag 300
ctcagtctcc aaatcttctg agcctgacaa acatatctga tatatttgc atcagtcc 360
tgtccattgc aagagccagt aacatagatg cagggaaagga caagctgggtt ccaggccaag 420
tcttactggt acctgttaact tgcgggtgcg ccggaaacca ctcttctgccc aatacctc 480
acccaaatcca gctagggtat agctacgact ttgttgcac cactttatata gagaacctta 540
caaattggaa tatagtacaa gttcaaaacc cagggtaaa tccatatttgc ttgccagagc 600
gcgtcaaaagt agtattccct ttattctgca ggtgcccttc aaagaaccag ttgaacaaag 660
ggattcagta tctgattact tatgtgtggc agcccaatga caatgtttcc cttgtgagtg 720
ccaaagttgg tgcattccccca gcgacatata tgactgaaaa ccgctacggc caagacttca 780
ctgctgcaac caacccatcca attttgcattcc cagtgcacaca gttgccagag cttactcaac 840
cttcttcaaa tggaaaggaaa agcagcattc atcttctggc tatacttggc attaccctgg 900
gatgcacggtt gctaactgca gttttacccg ggaccctcgat atatgtatac tgccgcagaa 960
agaaggctct gaataggact gcttcatcag ctgagactgc tgataaaacta ctttctggag 1020
tttcaggcta tgtaagcaag ccaaacgtgt atgaaatcga cgagataatg gaagctacga 1080
aggatttcag cgatgagtc aagggtgggg aatcagtgtc caaggccaac atagaaggc 1140
gggttgtagc ggtaaagaaa atcaaggaag gtggccaa tgaggaactg aaaattctgc 1200
agaaggtaaa tcatggaaat ctgggtgaaac taatgggtgt ctcctcaggc tatgtggaa 1260
actgtttctt ggttatgaa tatgctgaaa atgggtctct tgctgagtg cttttctcca 1320
agtcttcagg aaccccaaacc tcccttacat ggtctcaaag gataagcata gcagtgatg 1380
ttgctgtggg tctgcaatac atgcattgaaac atacccatcc aagaataata cacaggac 1440
tcacaacaag taatatcattt ctcgactcga acttcaaggc caagatagcg aatttcgcca 1500
tggccagaac ttcgaccaac cccatgatgc caaaaatcga tgtcttcgct ttcgggggtgc 1560
ttctgtataga gttgctcacc ggaaggaaaag ccatgacaac caaggagaac ggcgaggtgg 1620
ttatgctgtg gaaggatatg tggagatct ttgacataga agagaataga gaggagagga 1680

tcagaaaaatg gatggatcct aatttagaga gcttttatca tataaaaaat gctctcagct	1740
tggcatcctt agcagtgaat tgcacagctg ataagtcttt gtctcgaccc tccatggctg	1800
aaattgttct tagcctctcc tttctcaactc aacaatcatc taaccccaaca ttagagagat	1860
ccttgacttc ttctgggta gatgtagaag atgatgctca tattgtgact tccattactg	1920
cacgttaagc aagggaaaggt aattcagttt ctcataat tgcataagat gcactttgtt	1980
tgcgtggta ctattacatt tttaactagc tatttgccta ttctctgtt tttatggc	2040
agacactgga attgaatatc atatgatgga ggagttgtct gttataatcat gtgctaataa	2100
caaattcagg caagatagtt aattgcattt gaaatacata ttctgctca gagatggta	2160
acatccatgc tccgaagctc atattaagtg tggtagctat ttcttttca tcttttggg	2220
gtgaatgcgt gttcatgtaa ctcgtaaaggt gttatataatt acagaagtcg tatacgtcgt	2280
tccaaaaaaaaaa aa	2292

<210> 7
 <211> 3536
 <212> DNA
 <213> Lotus japonicus GIFU

<400> 7	
ggacatgaga ttgaagctcc aaaattagct ctttttctg atgaatactt aatgcattgt	60
tgtattcaact tgattaagtg ctagaaatca tctttgcattt atcatagatt aatgaattt	120
ccagttggtg tgtggagagc tattttgtt tgctgacatc tgcaatttgc agggcatcta	180
atgattgtca tttcttaat tattattggt tgttccgtt tcttaatatt tctgtttaa	240
tcttgcaggt catacaaatt aaaatactag ccaccaccca agacatacta aatggggtag	300
tagagggaaag ggtaaaggctg ataaggatga cttttatttcc tataaaattt aggagaattt	360
gagcttaagt ggcaaggcaa acgacattac tatacgaatt ggctttgtac cagaaacagg	420
gaacaaataa tattttacaa ataagctatt atcatgtcag ctcatttgcattt caactttgtat	480
ttgattaaaa attaaatgaa gttgaatttg ttgagctgct ttattatata tgccactgga	540
tgtttccgca ttctaaagtgc atgtttgaaa acatttctac aattgattac gaaggaaaaaa	600
ttaatcatgg agagaagctt atgtgcgttag cttctgtatt tctgaatttgc ttctatctgt	660
acagtagcat ttagataatg aatgatcttgc ttctcgctt agcataac caatctctac	720
ccttttaaaa ttgcaagaat tataagtcattt gcattgaccc aaatccttctt gtggttatgc	780
cccttaaaaaa tccggcaaga catcaagtta gttggtcattt agggttccac cagctagctg	840
acaccttgcata caacaactgg ccgtcctaaa gttggtaag cattacaata ctaaatgcctt	900

tttttattata ttttgcgcat ggttatatac ctaagtagga tttgtccaca gtttcttga 960
ttcggaaagg aaaaaatatt tagtgacac tgacagaagc agatttata tacatatatt 1020
atgaaatgac tcctacatga gatacacgaa tctcatcccc atgagttgca gtttgacaga 1080
gtacacactt atcaacttgc tggaatatag gaaagtctaa ccaatgtatgt cgatccgtat 1140
tgccttaatt ttggtaaatt tagtattaca tgatcattat tgatatacta aaccacagga 1200
tattttattg acaatgtgaa tggccatata tttcaacaat gctgattccc tctgataaag 1260
aacaagttcc ttttctcttt ccctgttaac tatcatttgt tccccacttc acaaacatgg 1320
ctgtcttctt tcttacctct ggctctctga gtctttct tgcaactcacg ttgctttca 1380
ctaacatcgc cgctcgatca gaaaagatta gcggcccaga ctttcatgc cctgttgact 1440
caccccttc ttgtgaaaca tatgtgacat acacagctca gtctccaaat cttctgagcc 1500
tgacaaacat atctgatata tttgatatac gtcccttgc cattgcaaga gccagtaaca 1560
tagatgcagg gaaggacaag ctggttccag gccaagtctt actggtagct gtaacttgcg 1620
gttgcgcgg aaaccactct tctgccaata cctcctacca aatccagcta ggtgatagct 1680
acgactttgt tgcaaccact ttatatgaga accttacaaa ttggaatata gtacaagctt 1740
caaaccagg ggtaaatcca tatttgc ttagcgcgt caaaagtagta ttccctt 1800
tctgcaggta cccttcaaag aaccagttga acaaaggat tcagtatctg attacttatg 1860
tgtggaagcc caatgacaat gttcccttgc tgagtgc 1920
acatatttgc taaaaaccgc tacggtaag acttcactgc tgcaaccaac ctcccaattt 1980
tgatcccagt gacacagttg ccagagctta ctcaaccttc ttcaaatgg 2040
gcattcatct tctggtaat ctgggttata ccctggatg cacgttgctt actgcagttt 2100
taaccggac cctcgatata gtatactgccc gcagaaagaa ggctctgaat aggactgctt 2160
catcagctga gactgctgat aaactacttt ctggagttc aggctatgta agcaagccaa 2220
acgtgtatga aatcgacgag ataatggaag ctacgaagga tttcagcgat gagtgcaagg 2280
ttggggatc agtgtacaag gccaacatag aaggtcggt tgtagcgta aagaaaatca 2340
aggaagggtgg tgccaaatgag gaactgaaaa ttctgcagaa ggtaaatcat ggaaatctgg 2400
tgaaaactaat ggggtctcc tcaggctatg atggaaactg tttcttggtt tatgaatatg 2460
ctgaaaatgg gtctcttgct gagtggtgt tctccaaatgc ttcaggaaacc ccaaactccc 2520
ttacatggtc tcaaaggata agcatagcag tggatgttgc tgggtctg caatacatgc 2580
atgaacatac ctatccaaga ataatacaca gggacatcac aacaagtaat atccttctcg 2640

actcgaactt caaggccaag atagcgaatt tcgccatggc cagaacttcg accaacccta 2700
tcatgcctaa aatcgatgtc ttgcgtttcg gggtgcttct gatagagttg ctcaccggaa 2760
ggaaagccat gacaaccaag gagaacggcg aggtggttat gctgtggaag gatatgtggg 2820
agatcttga catagaagag aatagagagg agaggatcg aaaatggatg gatcctaatt 2880
tagagagctt ttatcatata gataatgctc tcagcttggc atccttagca gtgaattgca 2940
cagctgataa gtctttgtct cgaccctcca tggctgaaat tggctttagc ctctccttcc 3000
tcactcaaca atcatctaac cccacattag agagatcctt gacttcttct gggtagatg 3060
tagaagatga tgctcatatt gtgacttcca ttactgcacg ttaagcaagg gaaggtaatt 3120
cagtttctca tcaaattgat caagatgcac tttgttgcg tggttactat tacatttta 3180
actagctatt tgcttatttc tctgtattta tttgtcagac actggattt aatatcatat 3240
gatggaggag ttgtctgtta atacatgtgc taataacaaa ttcaggcaag atagttaatt 3300
gcatttgaaa tacatatttc tgctcagaga tggtaacat ccatgctccg aagctcatat 3360
taagtgtggt agcttatttc tttcatctt tttgggtga atgcgtgttc atgtaactcg 3420
taagggttta tatattacag aagtcgtata cgtcgttcca ataattgatc aaggtaacctg 3480
tctattcgt aaaaaaaagcc aagtaccaac attagttgac tcgtttagag tggtgc 3536

<210> 8
<211> 595
<212> PRT
<213> Lotus japonicus

<400> 8

Met Ala Val Phe Phe Leu Thr Ser Gly Ser Leu Ser Leu Phe Leu Ala
1 5 10 15

Leu Thr Leu Leu Phe Thr Asn Ile Ala Ala Arg Ser Glu Lys Ile Ser
20 25 30

Gly Pro Asp Phe Ser Cys Pro Val Asp Ser Pro Pro Ser Cys Glu Thr
35 40 45

Tyr Val Thr Tyr Thr Ala Gln Ser Pro Asn Leu Leu Ser Leu Thr Asn
50 55 60

Ile Ser Asp Ile Phe Asp Ile Ser Pro Leu Ser Ile Ala Arg Ala Ser
65 70 75 80

Asn Ile Asp Ala Gly Lys Asp Lys Leu Val Pro Gly Gln Val Leu Leu

85

90

95

Val Pro Val Thr Cys Gly Cys Ala Gly Asn His Ser Ser Ala Asn Thr
100 105 110

Ser Tyr Gln Ile Gln Leu Gly Asp Ser Tyr Asp Phe Val Ala Thr Thr
115 120 125

Leu Tyr Glu Asn Leu Thr Asn Trp Asn Ile Val Gln Ala Ser Asn Pro
130 135 140

Gly Val Asn Pro Tyr Leu Leu Pro Glu Arg Val Lys Val Val Phe Pro
145 150 155 160

Leu Phe Cys Arg Cys Pro Ser Lys Asn Gln Leu Asn Lys Gly Ile Gln
165 170 175

Tyr Leu Ile Thr Tyr Val Trp Lys Pro Asn Asp Asn Val Ser Leu Val
180 185 190

Ser Ala Lys Phe Gly Ala Ser Pro Ala Asp Ile Leu Thr Glu Asn Arg
195 200 205

Tyr Gly Gln Asp Phe Thr Ala Ala Thr Asn Leu Pro Ile Leu Ile Pro
210 215 220

Val Thr Gln Leu Pro Glu Leu Thr Gln Pro Ser Ser Asn Gly Arg Lys
225 230 235 240

Ser Ser Ile His Leu Leu Val Ile Leu Gly Ile Thr Leu Gly Cys Thr
245 250 255

Leu Leu Thr Ala Val Leu Thr Gly Thr Leu Val Tyr Val Tyr Cys Arg
260 265 270

Arg Lys Lys Ala Leu Asn Arg Thr Ala Ser Ser Ala Glu Thr Ala Asp
275 280 285

Lys Leu Leu Ser Gly Val Ser Gly Tyr Val Ser Lys Pro Asn Val Tyr
290 295 300

Glu Ile Asp Glu Ile Met Glu Ala Thr Lys Asp Phe Ser Asp Glu Cys
305 310 315 320

Lys Val Gly Glu Ser Val Tyr Lys Ala Asn Ile Glu Gly Arg Val Val
325 330 335

Ala Val Lys Lys Ile Lys Glu Gly Gly Ala Asn Glu Glu Leu Lys Ile
340 345 350

Leu Gln Lys Val Asn His Gly Asn Leu Val Lys Leu Met Gly Val Ser
355 360 365

Ser Gly Tyr Asp Gly Asn Cys Phe Leu Val Tyr Glu Tyr Ala Glu Asn
370 375 380

Gly Ser Leu Ala Glu Trp Leu Phe Ser Lys Ser Ser Gly Thr Pro Asn
385 390 395 400

Ser Leu Thr Trp Ser Gln Arg Ile Ser Ile Ala Val Asp Val Ala Val
405 410 415

Gly Leu Gln Tyr Met His Glu His Thr Tyr Pro Arg Ile Ile His Arg
420 425 430

Asp Ile Thr Thr Ser Asn Ile Leu Leu Asp Ser Asn Phe Lys Ala Lys
435 440 445

Ile Ala Asn Phe Ala Met Ala Arg Thr Ser Thr Asn Pro Met Met Pro
450 455 460

Lys Ile Asp Val Phe Ala Phe Gly Val Leu Leu Ile Glu Leu Leu Thr
465 470 475 480

Gly Arg Lys Ala Met Thr Thr Lys Glu Asn Gly Glu Val Val Met Leu
485 490 495

Trp Lys Asp Met Trp Glu Ile Phe Asp Ile Glu Glu Asn Arg Glu Glu
500 505 510

Arg Ile Arg Lys Trp Met Asp Pro Asn Leu Glu Ser Phe Tyr His Ile
515 520 525

Asp Asn Ala Leu Ser Leu Ala Ser Leu Ala Val Asn Cys Thr Ala Asp
530 535 540

Lys Ser Leu Ser Arg Pro Ser Met Ala Glu Ile Val Leu Ser Leu Ser
545 550 555 560

Phe Leu Thr Gln Gln Ser Ser Asn Pro Thr Leu Glu Arg Ser Leu Thr
565 570 575

Ser Ser Gly Leu Asp Val Glu Asp Asp Ala His Ile Val Thr Ser Ile
580 585 590

Thr Ala Arg
595

<210> 9
<211> 23
<212> DNA
<213> Pisum sativum

<400> 9
atgtctgcct tctttcttcc ttc

23

<210> 10
<211> 23
<212> DNA
<213> Pisum sativum

<400> 10
ccacacataa gtaatmagat act

23

<210> 11
<211> 3800
<212> DNA
<213> Pisum sativum

<400> 11
gtgggctata tgattggtgc gtacttcacc ttgcataaaa tatcagcaca aagtatatca 60
agtaaaaaac aatacctaaa ttcccttaacc tatgatattc ttttggaga ggttgcaaaa 120
aagttgttag ttgcagttat tattttagtt ttgaaaatgt attgttggcc aaacatttagt 180
tgatactcag gaactagctc ttgttctgat ggatacttaa tgcttcgtta tatatttcta 240
ttcaacttgggt caagtgttag aaatcatctt ggcacaatca caggatgaat aaacctctgg 300
ttgaaagctt cattcagtcg ttgcgtgatt tctgcaactt gaggggaatc taatgatttt 360
tatttattat tattgcgttt gcttactgca attatcaatt ccttttaatt tttttacaaa 420
acaagttgggt tacaagatct cttaatata ttgttatcag ttatcagttt cttttatgta 480
agaagggttt ctctatacgg aactataaag actaattcctt caaatcgggt gggacaacaa 540
aagcggcaaa gttgttcatg aagaattttt gcaactgtgtt attcttatca agtacagaaa 600
gccacacactca agcaaaaaag tggatggtaa gaacgacatc ttattctatt ttatttagta 660

ggagaagtca agcttatgtg gcgatgtaaa tgtcattct atccaaacta tctttgtact 720
agaaataggg aacatataaa ttatggagag tttgttaagg tttttaata tattaaaacc 780
attgtAACGG gaagtgtcaa cattgttagc tggtcattgc ctgtatatta taatagcata 840
tatataatag acttggcctt tgtaaactt taaaccatat ctttgtgag tctaccctt 900
aaaaatatgg taaaggcatc aagttagata gtcttaggt accagccagc tagctgacat 960
tgtgttaagga catattggat tacaacta tattattt accatctta ttatattctg 1020
cgcatgattt catacttaat ttggattgt ccagtgtcta agattgaaa aggaaaaata 1080
gtagaactaa tgacagagac agaagcatat attttaata tcaaaccaaa agatatgtcc 1140
aaataagaga taaatataaa gtttgaggta taacaataag tcttggtgt tacttgccat 1200
aagaaactct ctttctctt ccccataact tgcattctt cacaattca caacaatggc 1260
tatcttcttt ctcccttcta gttctcatgc ctttttctt gcactcatgt ttttgcac 1320
taatatttca gctcaaccat tacaactca g tggacaaaac ttttcatgcc cgggtgattc 1380
acctccctca tgtgaaacct atgtgacata ctgtgcgg tctccaaact ttttgaggct 1440
aactaacata tcagatataat ttgatatgag tccttatacc attgcaaaag ccagtaacat 1500
agaagatgag gacaagaagc tgggtgaagg ccaagtctta ctcatacctg taacttgtgg 1560
ttgcactaga aatcgctatt tcgcgaattt cacgtacaca atcaagctag gtgacaacta 1620
tttcatagtt tcaaccactt cataccagaa tcttacaaat tatgtggaaa tggaaaattt 1680
caaccctaat ctaagtccaa atctattgcc accagaaatc aaagttgtg tccctttatt 1740
ctgcaaatgc ccctcgaaga atcagttgag caaaggaata aagcatctga ttacttatgt 1800
gtggcaggct aatgacaatg ttacccgtgt aagttccaag tttggtgcat cacaagtgg 1860
tatgtttact gaaaacaatc aaaacttcac tgcttcaacc aacgttccga ttttgatccc 1920
tgtgacaaag ttaccggtaa ttgatcaacc atcttcaaattt ggaagaaaaa acagcactca 1980
aaaacctgct tttataattt gtagatgcctt aggtatgtctt ttttgcgtt tagtttaac 2040
actatcactt gtttatgtat attgtctgaa aatgaagaga ttgaatagga gtacttcattt 2100
ggcggagact gcggataagt tactttcagg tggttcgggt tatgtaaagca agccaaacat 2160
gtatgaaatg gatgcgatca tggaaagctac aatgaacctg agtggaaattt gtaagattgg 2220
tgaatccgtt tacaaggcta atatagatgg tagatgtttt gtagtggaaa aaatcaagaa 2280
agatgcttctt gaggagctga aaattttgcgaa gaaggtaat catggaaatc ttgtgaaact 2340
tatgggtgtg tcttccgaca acgacggaaa ctgttccctt gtttacgagt atgctgaaaa 2400
tggatcactt gatgagtgg ttttctcaga gtcgtcgaaa acttcgaact cgggtggctc 2460

gcttacatgg tctcagagaa taacagtagc agtggatgtt gcagttggtt tgcaatacat	2520
gcatgaacat acttacccaa gaataatcca cagagacatc acaacaagta atatccttct	2580
ggattcaaac tttaaggcca agatagcgaa ttttcaatg gccagaactt caacaaattc	2640
catgatgccg aaaatcgatg ttttcgcttt tggggtggtt ctgattgagt tgcttaccgg	2700
caagaaaagcg ataacaacga tggaaaatgg cgaggtggtt attctgtgga aggatttctg	2760
gaagattttt gatctagaag ggaatagaga agagagctt agaaaaatggg tggatcctaa	2820
gctagagaat ttttattccta ttgataatgc tcttagtttgc gcttcttgg cagtgaatttgc	2880
tactgcagat aaatcattgt caagaccaag cattgcagaa attgttcttt gtcttctct	2940
tctcaatcaa tcatacatctg aaccaatgtt agaaagatcc ttgacatctg gtttagatgt	3000
tgaagctact catgttgtta cttctatagt agctcggttga tattcattca agtgaaggta	3060
acactgaatc aatgcttcag tttcttataat tcaagatggt tactttgtt agatgattat	3120
tgattacatc tttatgtgtg gaactatatg gtttattttaa ttaagggaat tggatcctaaa	3180
ttcatttttc catgttatttgc ttttacagca tgagttcgg taaagtgaat tgtaacctgc	3240
tattgaactc agaataattt cggttattat gtttagtcattc gacactttta agaaaagtat	3300
gtttgatgtt cgatataatgt ctgacaccaa cacaacactt acaactgtga ttatgtttaa	3360
tttggattttttt tttgtataa atcagtgttt catcatttgc ttatataaggta acaatttattc	3420
caaccatcct ttatataaggc cattctcttt atttttgtat acaatataag acctaagtgt	3480
gaatatttgc gcttaatggc agacatgaat tttgcaagaa aggatttggc agccttggc	3540
accataaaaa tggatgttgc agtcagctat aacttcttc tttttctttt tttttttggg	3600
atggatggg tattcatgtt tagctaaagg cacattttaa attaaaatct tttatataata	3660
tgcaaaaatgc ttctttgggtt ttcaataat ttttgcagg accgcttacc atcgatgggtt	3720
gagtttacaa taccacgtct atatatgtgg agaatcttc tcaaggatca agacttcgtt	3780
ggccagctgc taaaagacaa	3800

<210> 12
 <211> 2226
 <212> DNA
 <213> Pisum sativum

<400> 12	
ttttcttttc ctcataactt gcatttcttc acaatttcac aacaatggct atcttctttc	60
ttccttcttag ttctcatgcc ctttttcttg cactcatgtt ttttgcact aatatttcag	120
ctcaaccatt acaactcagt ggaacaaact tttcatgcc ggtggattca cctccttcat	180

gtgaaaccta tgtgacatac tttgctcggt ctccaaactt tttgagccta actaacatat 240
cagatatatt tgatatgagt cctttatcca ttgcaaaagc cagtaacata gaagatgagg 300
acaagaagct ggttgaaggc caagtcttac tcataacctgt aacttgtggt tgcactagaa 360
atcgctattt cgcgaaatttc acgtacacaa tcaagctagg tgacaactat ttcatagttt 420
caaccacttc ataccagaat cttacaaattt atgtgaaat ggaaaatttc aaccctaatac 480
taagtccaaa tctattgcca ccagaaatca aagttgtgt cccttattc tgcaaatgcc 540
cctcgaagaa tcagttgagc aaaggaataa agcatctgat tacttatgtg tggcaggcta 600
atgacaatgt tacccgtgta agtccaagt ttggtgcatc acaagtggat atgtttactg 660
aaaacaatca aaacttcact gcttcaacca atgttccgat tttgatccct gtgacaaagt 720
taccggtaat tgatcaacca tcttcaaattt gaagaaaaaa cagcactcaa aaacctgctt 780
ttataattgg tattagccta ggatgtgctt tttcgttgt agtttaaca ctatcacttg 840
tttatgtata ttgtctgaaa atgaagagat tgaataggag tacttcattt gcggagactg 900
cggataagtt actttcaggt gttcggggtt atgtaagcaa gccaaacaatg tatgaaatgg 960
atgcgatcat ggaagctaca atgaacctga gtgagaattt taagattggat gaatctgttt 1020
acaaggctaa tatagatggt agagtttag cagtaaaaaa aatcaagaaa gatgcttctg 1080
aggagctgaa aattctgcag aaggtaaatc atggaaatct tggaaactt atgggtgtgt 1140
cttccgacaa cgaaggaaac ttttccttgc ttacgagta tgctgaaaat ggatcacttg 1200
atgagtggtt gttctcagag ttgtcgaaaaa cttcgaactc ggtggtctcg cttacatggt 1260
ctcagagaat aacagtagca gtggatgtt cagttggttt gcaatacatg catgaacata 1320
cttacccaag aataatccac agagacatca caacaagtaa tttcattctg gattcaaact 1380
ttaaggccaa gatagcgaat ttttcaatgg ccagaacttc aacaaattcc atgatgccga 1440
aaatcgatgt ttgcgtttt ggggtggttc tgattgagtt gcttaccggc aagaaagcga 1500
taacaacgtt gaaaaatggc gaggtggttt ttctgtggaa ggatttctgg aagattttt 1560
atctagaagg gaatagagaa gagagcttaa gaaaatggat ggatcctaag cttagagaatt 1620
tttacccat tgataatgtt cttttttttt cttttttttt agtgaattttt actgcagata 1680
aatcattgtc aagaccaagc attgcagaaa ttgtttttt tttttttttt ctcaatcaat 1740
catcatctga accaatgtta gaaagatcct tgacatctgg ttttagatgtt gaagctactc 1800
atgttggtttac ttctatagta gctcggttatttcaaa gtgaaggttta cactaaatca 1860
atgcttcagt ttcttatattt caagatggttt actttttttt ggtgatttattt gattacatct 1920

ttatgtgtgg aactatatgg ttatTTTaat taaggaaatt agtctaaatt tcattttcc 1980
atgttattct ttaaagcacg agtttcggta aagtgaattg taacctgtta ttgagctcat 2040
aataatttca gttattatgt tagtcatcga cacttctaaa aaagtatgtc tgatgttcga 2100
tatgtgtctg acaccaacac aaccctgacc actgtgatta cgTTTaaattt gtttatttt 2160
gtgataaATC agtgtttcat catttgatta ttaaggTaca attattccaa ccATCCTTT 2220
aaaaaa 2226

<210> 13
<211> 1968
<212> DNA
<213> *Pisum sativum*

<400> 13
ctgcatttc ttcacaattt cacaacaatg gctatTTCT ttCTTCCttc tagTTCTcat 60
gcCCTTTTC ttgcactcat gTTTTTgtc actaatattt cagCTcaacc attacaactc 120
agtggAACAA actTTTcatg cccggTggat tcacCCttt catgtgaaac ctatgtgaca 180
tacTTTgTtc ggtctccaaa cTTTTgagc ctaactaaca tatcagatATt atttgatATg 240
agtCCTTTat ccattgcaaa agccagtaac atagaagatg aggacaagaa gctggTTgaa 300
ggCCAAGTCT tactcatacc tGtaacttGt ggtTgcacta gaaatcgcta ttTcgCgaat 360
ttcacgtaca caatcaagct aggtgacaac tatttcatAG ttcaaccac ttcataccAG 420
aatCTTacaA attatgtgga aatggAAAat ttcaacccta atctaagtcc aaatctattg 480
ccaccagAAA tcaaAGTTgt tGtccCTtta ttctgcaaat gcccctcgaa gaatcagttg 540
agcaaaggAA taaAGcatct gattacttAT gtgtggcagg ctaatgacAA tgTTaccCgt 600
gtaagttcca agTTTggTgc atcacaAGtg gatATgttA ctgAAAacAA tcaaAAacttc 660
actgCTTcaa ccaacgTTcc gatTTTgatc cctgtgacAA agttaccggT aattgatcaa 720
ccatCTTcaa atggaagAAA aaacagcact caaaaacctg cTTTataat tggTattAGc 780
ctaggatgtg cTTTTTcgt tGtagTTtta acactatcac ttgtttatgt atattgtctg 840
aaaatgaaga gattgaatAG gagTacttca ttggcggaga ctgcggataa gttactttca 900
ggTgTTTcgG gttatgtAAG caagccAAca atgtatgAAA tggatgcgt catggAAGct 960
acaatgaacc ttagtgagAA ttgtAAGatt ggtgaatccg ttacaaggc taatataGat 1020
ggtagagTTT tagcagtgAA aaaaatcaag aaagatgctt ctgaggagct gaaaattttg 1080
cagaaggtaa atcatggAAA tCTTgtgAAA cttatgggtg tGtCTTCCGA caacgacgga 1140
aactgtttcc ttgtttacga gtatgctgAA aatggatcac ttgtatgagtg gttgtctca 1200

gagtcgtcga aaacttcgaa ctcggggc tcgcttacat ggtctcagag aataacagta 1260
gcagtggatg ttgcagttgg tttgcaatac atgcataac atacttaccc aagaataatc 1320
cacagagaca tcacaacaag taatatcctt ctggattcaa actttaaggc caagatagcg 1380
aattttcaa tggccagaac ttcaacaaat tccatgatgc cgaaaatcga tgtttcgct 1440
tttgggtgg ttctgattga gttgcttacc ggcaagaaag cgataacaac gatggaaaat 1500
ggcgaggtgg ttattctgtg gaaggatttc tggaaagattt ttgatctaga agggaaataga 1560
gaagagagct taagaaaatg gatggatcct aagctagaga attttatcc tattgataat 1620
gctcttagtt tggcttctt ggcagtgaat tgtactgcag ataaatcatt gtcaagacca 1680
agcattgcag aaattgttct ttgtcttct cttctcaatc aatcatcatc tgaaccaatg 1740
ttagaaagat ctttgacatc tggtttagat gttgaagcta ctcatgttgc tacttctata 1800
gtagctcggt gatattcatt caagtgaagg taacactgaa tcaatgcttc agtttcttat 1860
attcaagatg gttactttgt ttagatgatt attgattaca tctttatgtg tggaaactata 1920
tggttatccc aattaaggaa attgttctaa aattcatttt tccatgtt 1968

<210> 14
<211> 1938
<212> DNA
<213> *Pisum sativum*

<400> 14
tcttcacaat ttcacaacaa tggctatctt ctttcttct tctagttctc atgccccttt 60
tctgcactc atgttttttgc tcaactaatat ttcagctcaa ccattacaac tcagtggaac 120
aaactttca tgcccgggtgg attcacctcc ttcatgtgaa acctatgtga catactttgc 180
tcggtctcca aacttttga gcctaactaa catatcagat atatttgcata tgagtccctt 240
atccattgca aaagccagta acatagaaga tgaggacaag aagctggttg aaggccaagt 300
cttactcata cctgtaactt gtggttgcac tagaaatcgc tatttcgcga atttcacgta 360
cacaatcaag cttagtgaca actatttcat agttcaacc acttcataacc agaatcttac 420
aaattatgtg gaaatggaaa atttcaaccc taatctaagt ccaaatctat tgccaccaga 480
aatcaaaagtt gttgtccctt tattctgcaa atgcccctcg aagaatcagt tgagcaaagg 540
aataaagcat ctgattactt atgtgtggca ggctaattgac aatgttaccc gtgttaagttc 600
caagtttgggt gcatcacaag tggatatgtt tactgaaaac aatcaaaact tcactgcttc 660
aaccatgtt ccgattttga tccctgtgac aaagttaccg gtaattgatc aaccatcttc 720
aaatggaaga aaaaacagca ctcaaaaacc tgctttata attggatttgcgcctaggatg 780

tgcttttc gttgtagtt taacactatc acttgttat gtatattgtc tgaaaatgaa 840
gagattgaat aggagtactt cattggcga gactgcggat aagttactt caggtgttc 900
gggttatgta agcaagccaa caatgtatga aatggatgcg atcatggaag ctacaatgaa 960
cctgagttag aattgtaga ttggtgaatc tggttacaag gctaatacg atggtagat 1020
tttagcagtg aaaaaaatca agaaagatgc ttctgaggag ctgaaaattc tgcagaaggt 1080
aaatcatgga aatcttgcg aacttatggg tggtcttcc gacaacgaa gaaactgtt 1140
ccttgcgttac gagtatgctg aaaatggatc acttgcgttag tggttgcgtc cagagttgtc 1200
gaaaacttcg aactcggtgg tctcgcttac atggtctcag agaataacag tagcagtgg 1260
tggtgcagtt gggttgcaat acatgcgtga acataacttac ccaagaataa tccacagaga 1320
catcacaaca agtaatatcc ttctggattc aaactttaag gccaagatag cgaatttttc 1380
aatggccaga acttcaacaa attccatgtat gcccggaaatc gatgtttcg cttttgggt 1440
ggttctgatt gagttgccta cggcaagaa agcgataaca acgatggaaa atggcgaggt 1500
ggttattctg tggaggatt tctggaaatgat ttttgcata gaaggaaata gagaagagag 1560
cttaagaaaa tggatggatc ctaagctaga gaattttat cctattgata atgctcttag 1620
tttggcttct ttggcagtga attgtactgc agataaatca ttgtcaagac caagcattgc 1680
agaaaattgtt ctttgcctt ctcttctcaa tcaatcatca tctgaaccaa tgtagaaag 1740
atccttgaca tctggtttag atgttgaagc tactcatgtt gttacttcta tagtagctcg 1800
ttgatattca ttcaagtgaa ggtaacacta aatcaatgct tcagtttctt atattcaaga 1860
tggttacttt gtttaggtga ttattgatta catctttag tggaaacta tatggttatt 1920
ttaattaagg gaatttagt 1938

<210> 15
<211> 594
<212> PRT
<213> Pisum sativum

<400> 15

Met Ala Ile Phe Phe Leu Pro Ser Ser Ser His Ala Leu Phe Leu Ala
1 5 10 15

Leu Met Phe Phe Val Thr Asn Ile Ser Ala Gln Pro Leu Gln Leu Ser
20 25 30

Gly Thr Asn Phe Ser Cys Pro Val Asp Ser Pro Pro Ser Cys Glu Thr
35 40 45

Tyr Val Thr Tyr Phe Ala Arg Ser Pro Asn Phe Leu Ser Leu Thr Asn
50 55 60

Ile Ser Asp Ile Phe Asp Met Ser Pro Leu Ser Ile Ala Lys Ala Ser
65 70 75 80

Asn Ile Glu Asp Glu Asp Lys Lys Leu Val Glu Gly Gln Val Leu Leu
85 90 95

Ile Pro Val Thr Cys Gly Cys Thr Arg Asn Arg Tyr Phe Ala Asn Phe
100 105 110

Thr Tyr Thr Ile Lys Leu Gly Asp Asn Tyr Phe Ile Val Ser Thr Thr
115 120 125

Ser Tyr Gln Asn Leu Thr Asn Tyr Val Glu Met Glu Asn Phe Asn Pro
130 135 140

Asn Leu Ser Pro Asn Leu Leu Pro Pro Glu Ile Lys Val Val Val Pro
145 150 155 160

Leu Phe Cys Lys Cys Pro Ser Lys Asn Gln Leu Ser Lys Gly Ile Lys
165 170 175

His Leu Ile Thr Tyr Val Trp Gln Ala Asn Asp Asn Val Thr Arg Val
180 185 190

Ser Ser Lys Phe Gly Ala Ser Gln Val Asp Met Phe Thr Glu Asn Asn
195 200 205

Gln Asn Phe Thr Ala Ser Thr Asn Val Pro Ile Leu Ile Pro Val Thr
210 215 220

Lys Leu Pro Val Ile Asp Gln Pro Ser Ser Asn Gly Arg Lys Asn Ser
225 230 235 240

Thr Gln Lys Pro Ala Phe Ile Ile Gly Ile Ser Leu Gly Cys Ala Phe
245 250 255

Phe Val Val Val Leu Thr Leu Ser Leu Val Tyr Val Tyr Cys Leu Lys
260 265 270

Met Lys Arg Leu Asn Arg Ser Thr Ser Leu Ala Glu Thr Ala Asp Lys
275 280 285

Leu Leu Ser Gly Val Ser Gly Tyr Val Ser Lys Pro Thr Met Tyr Glu
290 295 300

Met Asp Ala Ile Met Glu Ala Thr Met Asn Leu Ser Glu Asn Cys Lys
305 310 315 320

Ile Gly Glu Ser Val Tyr Lys Ala Asn Ile Asp Gly Arg Val Leu Ala
325 330 335

Val Lys Lys Ile Lys Lys Asp Ala Ser Glu Glu Leu Lys Ile Leu Gln
340 345 350

Lys Val Asn His Gly Asn Leu Val Lys Leu Met Gly Val Ser Ser Asp
355 360 365

Asn Asp Gly Asn Cys Phe Leu Val Tyr Glu Tyr Ala Glu Asn Gly Ser
370 375 380

Leu Asp Glu Trp Leu Phe Ser Glu Ser Ser Lys Thr Ser Asn Ser Val
385 390 395 400

Val Ser Leu Thr Trp Ser Gln Arg Ile Thr Val Ala Val Asp Val Ala
405 410 415

Val Gly Leu Gln Tyr Met His Glu His Thr Tyr Pro Arg Ile Ile His
420 425 430

Arg Asp Ile Thr Thr Ser Asn Ile Leu Leu Asp Ser Asn Phe Lys Ala
435 440 445

Lys Ile Ala Asn Phe Ser Met Ala Arg Thr Ser Thr Asn Ser Met Met
450 455 460

Pro Lys Ile Asp Val Phe Ala Phe Gly Val Val Leu Ile Glu Leu Leu
465 470 475 480

Thr Gly Lys Lys Ala Ile Thr Thr Met Glu Asn Gly Glu Val Val Ile
485 490 495

Leu Trp Lys Asp Phe Trp Lys Ile Phe Asp Leu Glu Gly Asn Arg Glu
500 505 510

Glu Ser Leu Arg Lys Trp Met Asp Pro Lys Leu Glu Asn Phe Tyr Pro

515 520 525

Ile Asp Asn Ala Leu Ser Leu Ala Ser Leu Ala Val Asn Cys Thr Ala
530 535 540

Asp Lys Ser Leu Ser Arg Pro Ser Ile Ala Glu Ile Val Leu Cys Leu
545 550 555 560

Ser Leu Leu Asn Gln Ser Ser Ser Glu Pro Met Leu Glu Arg Ser Leu
565 570 575

Thr Ser Gly Leu Asp Val Glu Ala Thr His Val Val Thr Ser Ile Val
580 585 590

Ala Arg

<210> 16
<211> 19
<212> DNA
<213> Lotus japonicus

<400> 16
tgcatggca tggagaacc 19

<210> 17
<211> 20
<212> DNA
<213> Lotus japonicus

<400> 17
tttgctgtga cattatcagc 20

<210> 18
<211> 20
<212> DNA
<213> Lotus japonicus

<400> 18
ttgcagattt cacaactagg 20

<210> 19
<211> 21
<212> DNA
<213> Lotus japonicus

<400> 19
acttagaattc tgcaacttt c 21

<210> 20
<211> 21
<212> DNA
<213> *Lotus japonicus*

<400> 20
acttagaaatc tgcaactttg c

21

<210> 21
<211> 2205
<212> DNA
<213> *Lotus japonicus*

<400> 21
aagtgtgaca ttagtttcaa gagaaaaata aatgatcaaa acctggtaga gagtcctaga 60
aattcaatgt tctgatttct ttcattcatc tctgctgcca ttttgatttg cacaatgaag 120
ctaaaaactg gtctactttt gttttcatt ctttgctgg ggcatgttg tttccatgtg 180
gaatcaaaact gtctgaaggg gtgtgatcta gcttttagctt cctattatat cttgcctggt 240
gttttcatct tacaaaacat aacaaccttt atgcaatcag agattgtctc aagtaatgat 300
gccataacca gctacaacaa agacaaaatt ctaaatgata tcaacatcca atcctttcaa 360
agactcaaca ttccatttcc atgtgactgt attgggtggg agtttctagg gcatgttatt 420
gagtaactcag cttcaaaagg agacacttat gaaactattt ccaacctcta ctatgcaa 480
ttgacaacag ttgatctttt gaaaagggttc aacagctatg atccaaaaaaa catacctgtt 540
aatgccaagg ttaatgtcac tgtaattgt tcttgggata acagccaggt ttcaaaagat 600
tatggcttgc ttattaccta tcccatttagg cctggggata cactgcagga tattgcaa 660
cagagtagtc ttgatgcagg gttgatacag agtttcaacc caagtgtcaa tttcagcaa 720
gatagtggga tagtttcat tcctggaga tataaaaatg gagtctatgt tcccttgc 780
cacagaaccg caggtctagc tagtggtgca gctgttggta tatctattgc aggaacacctc 840
gtgcttgc ttacttagt ttgtatgtat gtttagatacc agaagaagga agaagagaaa 900
gctaaattgc caacagatata ttctatggcc ctttcaacac aagatgcctc tagtagtgca 960
gaatatgaaa cttctggatc cagtgccca gggactgcta gtgctacagg tcttactagc 1020
attatggtgg cgaaatcaat ggagttctca tatcaggaac tagcgaaggc tacaaataac 1080
tttagcttgg ataataaaaat tggtcaaggt ggatttggag ctgtcttatta tgcagaattg 1140
agaggcaaga aaacagcaat taagaagatg gatgtacaag catcaacaga atttctttgt 1200
gagttgaagg tcttaacaca tgttcaccac ttgaatctgg tgcgcttgat tggatactgc 1260
gttgaggat ctctattccct tgtttatgaa catattgaca atggaaactt aggccaatat 1320

ttgcatggtt caggtaaaga accattgcca tggtctagcc gagtacaaat agctctagat 1380
gcagcaagag gccttgaata cattcatgag cacactgtgc ctgtgtatat ccacgcgc 1440
gtgaaatctg caaacatatt gatagataag aacttgcgtg gaaagggtgc agatttggc 1500
ttgaccaagc ttattgaagt tggaaactcc acactacaaa ctcgtctggg gggacattt 1560
ggatacatgc ccccagaata tgctcaatat ggtgatattt ctccaaaaat agatgtat 1620
gcatttggag ttgttcttt tgaacttatt tctgcaaaga atgctgttct gaagacaggt 1680
gaatttagttg ctgaatcaaa gggccttgta gctttgttg aagaagcact taataagagt 1740
gatccttggat atgctttcg caaactggtg gatccttaggc ttggagaaaa ctatccaatt 1800
gattctgttc tcaagattgc acaacttaggg agagcttgta caagagataa tccactgcta 1860
agaccaagta tgagatctt agttgttgct cttatgaccc tttcatcaact tactgaggat 1920
tgtgatgatg aatcttccta cgaaagtcaa actctcataa atttactgtc tgtgagataa 1980
aggttctcca tgcaaattgca tggggtttat atatatctt tagtacaact aagcagacaa 2040
aaagtttgtt actttgaatg taaatcgagt caggggtttt acattttattt actccaatgt 2100
ttaattgcca aaaccatcaa aaagtccctag gccagacttc ctgttaattt atttagcaaa 2160
gttgcagatt ctaagttcag tttttttaaa aaaaaaaaaa aaaaa 2205

<210> 22
<211> 2210
<212> DNA
<213> Lotus japonicus

<400> 22
aagtgtgaca ttagttcaa gagaaaaata aatgatcaaa acctggtaga gagtcctaga 60
aattcaatgt tctgatttct ttcattcattc tctgctgcca ttttgatttg cacaatgaag 120
ctaaaaactg gtctactttt gttttcatt ctttgctgg ggcatgttg tttccatgtg 180
gaatcaaact gtctgaaggg gtgtgatcta gctttagctt cctattatat cttgcctgg 240
gttttcatct tacaaaacat aacaaccttt atgcaatcag agattgtctc aagtaatgt 300
gccataacca gctacaacaa agacaaaattt ctcaatgata tcaacatcca atccttcaa 360
agactcaaca ttccatttcc atgtgactgt attgggtggg agtttctagg gcatgtattt 420
gagtaactcag cttcaaaagg agacacttat gaaactattg ccaacctcta ctatgcaa 480
ttgacaacag ttgatctttt gaaaagggttc aacagctatg atccaaaaaaa catacctgtt 540
aatgccaagg ttaatgtcac tgtaattgt tcttgggaa acagccaggt ttcaaaagat 600
tatggcttgc ttattaccta tcccatcagg cctggggata cactgcagga tattgcaac 660

cagagtagtc ttgatgcagg gttgatacag agtttcaacc caagtgtcaa tttcagcaaa 720
gatagtggga tagcttcat tcctggaaga tataaaaatg gagtctatgt tcccttgtac 780
cacagaaccg caggtcttagc tagtggtgca gctgtggta tatctattgc aggaaccttc 840
gtgcttctgt tactagcatt ttgtatgtat gttagatacc agaagaagga agaagagaaa 900
gctaaattgc caacagatat ttctatggcc ctttcaacac aagatggtaa tgcctctagt 960
agtgcagaat atgaaacttc tggatccagt gggccagggta ctgctagtgc tacaggcttt 1020
actagcatta tggtggcgaa atcaatggag ttctcatatc aggaacttagc gaaggctaca 1080
aataacttta gcttggataa taaaattggt caaggtggat ttggagctgt ctattatgca 1140
gaattgagag gcaagaaaac agcaattaag aagatggatg tacaagcatac aacagaattt 1200
cttgtgagt tgaaggcttt aacacatgtt caccacttga atctggtgcg cttgattgga 1260
tactgcgttg agggatctct attccttgtt tatgaacata ttgacaatgg aaacttaggc 1320
caatatttgc atggttcagg taaagaacca ttgccatggt cttagccgagt acaaatacg 1380
ctagatgcag caagaggcct tgaatacatt catgagcaca ctgtgcctgt gtatatccat 1440
cgcgatgtga aatctgcaaa catattgata gataagaact tgcgtggaaa gttgcagat 1500
tttggcttga ccaagcttat tgaagttggg aactccacac tacaaactcg tctggtggga 1560
acatttggat acatgcffff agaatacgct caatatggtg atatttctcc aaaaatagat 1620
gtatatgcat ttggagttgt tcttttggaa cttatttctg caaagaatgc tgttctgaag 1680
acaggtgaat tagttgctga atcaaaggcc cttgtagctt tgtttgaaga agcacttaat 1740
aagagtgate cttgtgatgc tcttcgcaaa ctgggtggatc cttaggcttgg agaaaaactat 1800
ccaatttgatt ctgttctcaa gattgcacaa cttagggagag cttgtacaag agataatcca 1860
ctgctaagac caagtatgag atcttttagtt gttgctctta tgacccttc atcacttact 1920
gaggattgtg atgatgaatc ttccctacgaa agtcaaactc tcataaattt actgtctgt 1980
agataaaaggt tctccatgca aatgcatttt tgtttatataat atctttagt acaactaagc 2040
agacaaaaaag ttttgtactt tgaatgtaaa tcgagtcagg gtgtttacat tttattactc 2100
caatgtttaa ttgccaaaac cataaaaaag tcctaggcca gacttcctgt aattatattt 2160
agcaaaagttg cagattctaa gttcagtttt tttaaaaaaaa aaaaaaaaaa 2210

<210> 23
<211> 10253
<212> DNA
<213> Lotus japonicus Gifu

<220>

<221> exon
<222> (4172)..(4808)
<223>

<220>
<221> Intron
<222> (4809)..(5280)
<223>

<220>
<221> exon
<222> (5281)..(5314)
<223>

<220>
<221> Intron
<222> (5315)..(5561)
<223>

<220>
<221> exon
<222> (5562)..(5569)
<223>

<220>
<221> Intron
<222> (5570)..(5685)
<223>

<220>
<221> exon
<222> (5686)..(5838)
<223>

<220>
<221> Intron
<222> (5839)..(6475)
<223>

<220>
<221> exon
<222> (6476)..(6678)
<223>

<220>
<221> Intron
<222> (6679)..(7105)
<223>

<220>
<221> exon
<222> (7106)..(7195)
<223>

<220>
<221> Intron
<222> (7196)..(7933)
<223>

<220>
<221> exon
<222> (7934)..(8027)
<223>

<220>
<221> Intron
<222> (8028)..(8232)
<223>

<220>
<221> exon
<222> (8233)..(8384)
<223>

<220>
<221> Intron
<222> (8385)..(8471)
<223>

<220>
<221> exon
<222> (8472)..(8563)
<223>

<220>
<221> Intron
<222> (8564)..(9137)
<223>

<220>
<221> exon
<222> (9138)..(9275)
<223>

<220>
<221> Intron
<222> (9276)..(9403)
<223>

<220>
<221> exon
<222> (9404)..(9502)
<223>

<220>
<221> Intron
<222> (9503)..(9694)
<223>

<220>
<221> exon
<222> (9695)..(9859)
<223>

ttttccgcc atatcttaat tttgtttta taaaattaa aaagaatatg attgaatgtc 1260
aatgtaaatt tttttcacac agacaatgca tatccattaa ggtttgttag aattacactc 1320
ccccccattt ttatctaaaa tctacatccc accccatttt atatagaggc aaatttagtg 1380
acgaaaaata ttcttcatta ataatttagtt attatttaaa ttgttaatca ataattcaa 1440
aaaaaaattc aatataatcc aataaattta aaaatgaaaa catcacaatc ctcctcattc 1500
tctcaatcgc gtttacctc cgtaaattta caatgcaaat tatgcaatag cacacctgcc 1560
cgatttacaa accatatttc gaacatagtg aaacatgctt gtgtttcat atttggtgat 1620
aattcaattt taatcaaaat aatctcttta tacctccaat tttcaaaatt gggttgtagg 1680
ccaaaaaaagc aacacaaaatg ggtgaagaaa atagagaaac aaaattatga aaatatgaag 1740
tggatctgag gttttagag cccaacgagg cggtggagca tcgttttaa caaaatccaa 1800
caatatcttt aggggtgaaa tccaccaacc gagcgttcgc tcacaactta aaggggtgaa 1860
attcacaaga agtagttaa aggggtgaaa ttcacaagaa gtagttgaac aagtgacttt 1920
aggaatgtgc gattcacgtt ctggggttca ggtcgcaaca aaactttgag gcatggtggt 1980
gccatgtggt ttcacattgt gggacagtgg agccgtgtta aagggagtaa aggcttggtg 2040
gtggccgttt gtggtgaaaa atatgattt gагатагатг ggacgtggac ttaacagaca 2100
gагтggatgg tttttttta aatttaattc agtaaaatta tttttataaa ttaatgtatg 2160
atagtgtata tgcattaaat ttattnaaat ttttactaat tagtaattt gttttagtag 2220
tgacgaattt gttttgtca ctaaattttg ctttataaa aatggggtg gагtgtagat 2280
ttttaaaaag aatggggtg agtgcattc ttgcaaattct tgaggggggt gагtgtatTT 2340
tactcaactc ttaaaaaat taggaattaa ttagttgaa attataaaag ttttatttcat 2400
tgaataacat aacaattaa aggcaaaaaa atacaaaact tcattttata tgtatttcag 2460
aaaaattgcc tacttcaat tatgagaaac taaaattatg tttagttaa aatgagcata 2520
gattcaaaaaa ttaataaaata atatataatg caggatacat gcctatcaat taacatatcg 2580
tttgtccacg atgatgatct tattggagga tcaatatctt caaattaaca aagttatcac 2640
ttggctctta ttggcataa tgcataaaaa aaattgcaat tagtatcaaa tcaaactgaa 2700
atttgcaact atatgctgct ggtgtgtcg cgcaattcc tttttgattt ttatggaaat 2760
gaagtcaatg aagcaacagt ttcacaggcg tgctaaaaa taaaaaaatt ggaaatttga 2820
tgtttggtag gattatgaga ggacacaatg ggaggatgtt tcacaagctg cagacagggt 2880
tgccacttca gatgcaaagg attaaataaa caaagccaaag gtttgcattc aacaagattc 2940

catcgtcgtt ttgcttcctt taatcgatt aatcaaaagc acaccaagta aagcatcaat	3000
atataacatc caagaaatca caacatgata gttgctcgtc tcgtctatta actatgatgt	3060
caggagttcg atccccgctc atgtgaatgg aagacatttc gttgttagat gtttaccgtt	3120
taatgcaaat actcgccgtg agataataag tcattgtgt gggcgaatac cctaaaataa	3180
gaataaaatt aaatatacgca tccaaagttat tgcccaaata tataaacaat ggtattgtt	3240
acattattag gcataaaaagc agtaggtaag tgtattatat ttatthaatt ttttaaaatt	3300
ttgaaattaa ttaataattt ttaacataag taaaccattt ttagcaaaaa ctctacactt	3360
ctattacctt aacaagtaca ttttgatgg tacaccttaa caattaacaa gtcatatgat	3420
tgacaaacat attttatatg ctttacaattt tattctaaaa tcaaagttt tggaagaag	3480
ctcataaaaag tagttcctgg gtgttttttta gaatagagaa gttgatcatg ttagaaattt	3540
agtaaaaat gagttgaaag tgatttatgt ttgattatat ttatgagaaa aatgaattgt	3600
ctgatgtaat attgtaaaat ctaacaatttta attaagtacc acagaaacta gaatttata	3660
cttcaccttta gaattgattt tggagttaaa atcaattattt aaaggagcaa ttatcaaagg	3720
agacatccaa atacactagt taattttgac aatcaattct aacacttgca aatgtgttac	3780
caaacttact atcagtaagt gaactaatga ttcccaagtc aacttttggc ctagctagcc	3840
aaccgttact atgtcccttc cacaatacat tctccttggaa actgtcaagt gtcaactgca	3900
cccaaacatc cttgtttgtg atgaaaagat cgaaaacgtg tgcttatgaa tttacatgtt	3960
tacattcacc aaaaatcaaa agttacaccc tctatactt cacatatgtt tgagtcaactt	4020
tccatataaa atcccatagt ctattaatta tcagagtaag tggacatca gtttcaagag	4080
aaaaataaaat gatcaaaaacc tggtagagag tcctagaaat tcaatgttct gatttcttc	4140
attcatctct gctgccattt tgatttgac a atg aag cta aaa act ggt cta	4192
Met Lys Leu Lys Thr Gly Leu	
1 5	
ctt ttg ttt ttc att ctt ttg ctg ggg cat gtt tgt ttc cat gtg gaa	4240
Leu Leu Phe Phe Ile Leu Leu Leu Gly His Val Cys Phe His Val Glu	
10 15 20	
tca aac tgt ctg aag ggg tgt gat cta gct tta gct tcc tat tat atc	4288
Ser Asn Cys Leu Lys Gly Cys Asp Leu Ala Leu Ala Ser Tyr Tyr Ile	
25 30 35	
ttg cct ggt gtt ttc atc tta caa aac ata aca acc ttt atg caa tca	4336
Leu Pro Gly Val Phe Ile Leu Gln Asn Ile Thr Thr Phe Met Gln Ser	
40 45 50 55	
gag att gtc tca agt aat gat gcc ata acc agc tac aac aaa gac aaa	4384
Glu Ile Val Ser Ser Asn Asp Ala Ile Thr Ser Tyr Asn Lys Asp Lys	
60 65 70	

att ctc aat gat atc aac atc caa tcc ttt caa aga ctc aac att cca Ile Leu Asn Asp Ile Asn Ile Gln Ser Phe Gln Arg Leu Asn Ile Pro 75 80 85	4432
ttt cca tgt gac tgt att ggt ggt gag ttt cta ggg cat gta ttt gag Phe Pro Cys Asp Cys Ile Gly Gly Glu Phe Leu Gly His Val Phe Glu 90 95 100	4480
tac tca gct tca aaa gga gac act tat gaa act att gcc aac ctc tac Tyr Ser Ala Ser Lys Gly Asp Thr Tyr Glu Thr Ile Ala Asn Leu Tyr 105 110 115	4528
tat gca aat ttg aca aca gtt gat ctt ttg aaa agg ttc aac agc tat Tyr Ala Asn Leu Thr Thr Val Asp Leu Leu Lys Arg Phe Asn Ser Tyr 120 125 130 135	4576
gat cca aaa aac ata cct gtt aat gcc aag gtt aat gtc act gtt aat Asp Pro Lys Asn Ile Pro Val Asn Ala Lys Val Asn Val Thr Val Asn 140 145 150	4624
tgt tct tgt ggg aac agc cag gtt tca aaa gat tat ggc ttg ttt att Cys Ser Cys Gly Asn Ser Gln Val Ser Lys Asp Tyr Gly Leu Phe Ile 155 160 165	4672
acc tat ccc att agg cct ggg gat aca ctg cag gat att gca aac cag Thr Tyr Pro Ile Arg Pro Gly Asp Thr Leu Gln Asp Ile Ala Asn Gln 170 175 180	4720
agt agt ctt gat gca ggg ttg ata cag agt ttc aac cca agt gtc aat Ser Ser Leu Asp Ala Gly Leu Ile Gln Ser Phe Asn Pro Ser Val Asn 185 190 195	4768
ttc agc aaa gat agt ggg ata gct ttc att cct gga aga t gtatgttac Phe Ser Lys Asp Ser Gly Ile Ala Phe Ile Pro Gly Arg 200 205 210	4818
ctttttgttt taaatttttc cgctttgatt aaagtttatt attattagca tgattggatc	4878
aacttcttctt tcatcaaaat catttctgaa actcagaagc tactcacaca agcttcctgg	4938
tttcagaatc aattgttagta gggtttccaa acatgctttt ttatcaaaat caattacgta	4998
actcagaaac tactcacata agtttctct tagaattgat tctgttttta gaatcaattg	5058
taaaagggtt tacaaacatg cactctgcta gtgtgtgtc ttaaaaactat tcatggtaa	5118
attactcttc cattgtttctt acaataatac atgacaaggc atgtaactta ccccacctaa	5178
ttgaaaaatg gttgggtgtt attgttatat catttgtca atacatttga tataaacttt	5238
tatgaattta cctgaagttt tacttttctt tgaacttttc ag at aaa aat gga Tyr Lys Asn Gly 215	5291
gtc tat gtt ccc ttg tac cac ag gtgggttaact tcaattgcct actcatctt Val Tyr Val Pro Leu Tyr His Arg	5344

ttatgtatgaa tgatagcatg tttggatcaa cttctcttc accagaatta atccttaat	5404
tcagaactaa gaagctactc acataagctt tttcccgaa ttaattctgg cttcagaagc	5464
aattacactg aaagatttcc aaacatgctc taaatattgt ttcgtgcttg gttctatctt	5524
tttaacttcc atttattttt ccttttcat tttgcag a acc gca g gttggccct	5579
Thr Ala	
225	
ctaaatggc tctaggatg attattttt ccttgatgtt cacaaaaata tgagaacaca	5639
aaaaaaagagg atgcctctga gcttagctt acttctatgt aagcag gt cta gct	5693
Gly Leu Ala	
agt ggt gca gct gtt ggt ata tct att gca gga acc ttc gtg ctt ctg	5741
Ser Gly Ala Ala Val Gly Ile Ser Ile Ala Gly Thr Phe Val Leu Leu	
230 235 240 245	
tta cta gca ttt tgt atg tat gtt aga tac cag aag aag gaa gaa gag	5789
Leu Leu Ala Phe Cys Met Tyr Val Arg Tyr Gln Lys Lys Glu Glu Glu	
250 255 260	
aaa gct aaa ttg cca aca gat att tct atg gcc ctt tca aca caa gat g	5838
Lys Ala Lys Leu Pro Thr Asp Ile Ser Met Ala Leu Ser Thr Gln Asp	
265 270 275	
gtaatggat atttccaaat tcataattcct tctaaggctt aaccctctt agtccccctg	5898
gaaatgggtg aatgttggtg ctctaatttt tcatgtgttt aaatcagttt tatactaaga	5958
gtctgttggca caacaggttt ttgttttaa aacagaaaaa gccggaaatt tggttgatat	6018
gaaaagttt aaggaaattt ttatttttt gatatatcgg aaaattctt ttaagtgttc	6078
ctgttctcat tttctaaaac taaaatttca aaacatctcg gaggattttt cttcttgg	6138
ttagtttca attcacaggt ctttcagttt tgtaagcatc ttgttcaaat atagattttc	6198
ttttcttctt ttgaaaaaca tgcataaaa ttatttctga aaatagttt taaattttaga	6258
ggactgagaa gagaatcaa caagtcctaa ttttacatt ttcctgttta tcatttataa	6318
acttattacc tgatctaatt tcaggctaca ttttacatt tgcattaaaggc agaaaatttta	6378
cctgatccaa atgtttgagt tccattcaat ctggcacatt gatataattt gagaggat	6438
gacaacacta gctaactttt cttccctttt cttgaag cc tct agt agt gca gaa	6492
Ala Ser Ser Ser Ala Glu	
280	
tat gaa act tct gga tcc agt ggg cca ggg act gct agt gct aca ggt	6540
Tyr Glu Thr Ser Gly Ser Ser Gly Pro Gly Thr Ala Ser Ala Thr Gly	
285 290 295	
ctt act agc att atg gtg gcg aaa tca atg gag ttc tca tat cag gaa	6588
Leu Thr Ser Ile Met Val Ala Lys Ser Met Glu Phe Ser Tyr Gln Glu	
300 305 310 315	

cta gcg aag gct aca aat aac ttt agc ttg gat aat aaa att ggt caa	6636
Leu Ala Lys Ala Thr Asn Asn Phe Ser Leu Asp Asn Lys Ile Gly Gln	
320 325 330	
ggt gga ttt gga gct gtc tat tat gca gaa ttg aga ggc aag	6678
Gly Gly Phe Gly Ala Val Tyr Tyr Ala Glu Leu Arg Gly Lys	
335 340 345	
gtatgtgaccg tgggtctctt cagttctata acatagtgc tgggtggata caaagaggaa	6738
aaccacggtg aagccaaatt tgcgggtggac agacacaaaa gctaaaggaa gttgtcacca	6798
tgttttcaa ttgtgtatcc aaacttgcac aaaagaggat agaagtttct tacatttagag	6858
tagtagtgaa aagtttaat tttaaggctt tgggttcatt gtgaggaagc tatataaaac	6918
aactcaaaatc agtttagggc aaaaaattgt ttcatggaaa agaaagataa gagtaatgat	6978
tttacttaaa tggatattgt tcttaaagag gtggatggaa aagtttctgc tttttgtgcc	7038
acttttaggtt atccctttaa cttaactc ttccctggatt tcctctaattt caattttattc	7098
aatgcag aaa aca gca att aag aag atg gat gta caa gca tca aca gaa	7147
Lys Thr Ala Ile Lys Lys Met Asp Val Gln Ala Ser Thr Glu	
350 355	
ttt ctt tgt gag ttg aag gtc tta aca cat gtt cac cac ttg aat ctg	7195
Phe Leu Cys Glu Leu Lys Val Leu Thr His Val His His Leu Asn Leu	
360 365 370 375	
gtacaacatc cttaaacaacaa cttaaagcat tattatatct ttgggaagga aagattaata	7255
ttttatgtt tagtttgaag aatcatttagg ttcttacaaa acaaatatcc ttcatggttc	7315
tgtgaactga atagtcctat agttatccag caaaatttct gcagatccac atgatagtcc	7375
aacatggat ctgcattact agtgaagaa ctgtaaaac atttgttaact tcaattttct	7435
gtccttgaaa gtaacagacc atttagagca cactccccaa cattaatacc aaataaaagaa	7495
gaaaatcagc cctttcccg catgtgtggt tccactgtga aatatttcaa aatcacttgt	7555
gattagaagc tacaagtcta agtttctgag caaacgtgtc ttggattttg tgctaatcat	7615
aaagccaaat atgctattag ttaatgatta aaggcattat tagaaactcc tttatttcca	7675
attgccactg ttgatatgtt atttggattt ttcaaacagt ttctcctaacc aacagggttc	7735
agaaaaaaaaa ttagtattaa ttcttatcta tgattactta aagaagaaag tgctaaattc	7795
tttctggat ttcaatataa ctatatcata cactttcat ttaattttc taattttgga	7855
atctttgttt agcataaaca gctctaagta agttataatt cttattctgt atgtacctac	7915
tttctatgaa caacatag gtg cgc ttg att gga tac tgc gtt gag gga tct	7966
Val Arg Leu Ile Gly Tyr Cys Val Glu Gly Ser	
380 385	

ctt ttc ctt gtt tat gaa cat att gac aat gga aac tta ggc caa tat Leu Phe Leu Val Tyr Glu His Ile Asp Asn Gly Asn Leu Gly Gln Tyr 390 395 400	8014
ttg cat ggt tca g gtgagaacag gatgcagtga tatttttttgc ctgtgacatt Leu His Gly Ser 405	8067
atcagcatgt ttggatcaat ttctctttca ccagaattaa ttctgaaaca gagaagtagc ttctccacag aattgattct gacttcagag tcaatagtag aattatccg aaacatgcac ggcattatag tcaaacaatt aataatgatg atgacatgat ttcag gt aaa gaa cca Gly Lys Glu Pro 410	8127 8187 8243
ttg cca tgg tct agc cga gta caa ata gct cta gat gca gca aga ggc Leu Pro Trp Ser Ser Arg Val Gln Ile Ala Leu Asp Ala Ala Arg Gly 415 420 425	8291
ctt gaa tac att cat gag cac act gtg cct gtg tat atc cat cgc gat Leu Glu Tyr Ile His Glu His Thr Val Pro Val Tyr Ile His Arg Asp 430 435 440	8339
gtg aaa tct gca aac ata ttg ata gat aag aac ttg cgt gga aag Val Lys Ser Ala Asn Ile Leu Ile Asp Lys Asn Leu Arg Gly Lys 445 450 455	8384
gttgcattta ttaccaatct tcatgatcca aattctttca tttcttctt gagactttaa	8444
tcaaactgtg aaagttttta tgttcag gtt gca gat ttt ggc ttg acc aag ctt Val Ala Asp Phe Gly Leu Thr Lys Leu 460 465	8498
att gaa gtt ggg aac tcc aca cta caa act cgt ctg gtg gga aca ttt Ile Glu Val Gly Asn Ser Thr Leu Gln Thr Arg Leu Val Gly Thr Phe 470 475 480	8546
gga tac atg ccc cca ga gtatgatttt cttttgcgt tgattaaatg Gly Tyr Met Pro Pro Asp 485	8593
gtgttttgg ataaacagtt taatcaaaag ttgatggtaa taaacaccta tcgcataagt	8653
gtttattcat aaactatttt gagatgttta ttgagataaa gttaaaatat ctaatgagtt	8713
tagtgactta tgaaagtaag ctctcaacaa cttataagta gggataagg tatttacaat	8773
acataagctc taacaagcac ttagatacac acatttgcgt ttatctttca caataaatgc	8833
tcgtacaagt gtttgcgt gcttgcgt accttgcgt tgatttgcgt	8893
ttatttcac aagttgttca tattagcttta tgaataagag attatgcttataatatt	8953
attttcagct tatttcaata agttcatcaa atttgcgttataatatttgcgt	9013
agcgcttatt gctacaagtgc ttaattacg ctgtttaccc ataaacgtgt tcaatttagta	9073
aagtcaagtt cagtttcaa aacatatcat tgagtgaact tgtttaccc ggctttatgt	9133

caga t atg ctc aat atg gtg ata ttt ctc caa aaa tag atg tat atg		9180
Met Leu Asn Met Val Ile Phe Leu Gln Lys	Met Tyr Met	
490	495	500
cat ttg gag ttg ttc ttt ttg aac tta ttt ctg caa aga atg ctg ttc		9228
His Leu Glu Leu Phe Phe Leu Asn Leu Phe Leu Gln Arg Met Leu Phe		
505	510	515
tga aga cag gtg aat tag ttg ctg aat caa agg gcc ttg tag ctt tg		9275
Arg Gln Val Asn	Leu Leu Asn Gln Arg Ala Leu	Leu Cys
520	525	
gtgagtcac atgccccttc tctaacctta tttacaaacc aattactcac aatttcgaaa		9335
attttacatg tatatttcaa agctactcag cacaaatgca tttgccctta acttgctttg		9395
cattgcag t ttg aag aag cac tta ata aga gtg atc ctt gtg atg ctc		9443
Leu Lys Lys His Leu Ile Arg Val Ile Leu Val Met Leu		
535	540	
ttc gca aac tgg tgg atc cta ggc ttg gag aaa act atc caa ttg att		9491
Phe Ala Asn Trp Trp Ile Leu Gly Leu Glu Lys Thr Ile Gln Leu Ile		
545	550	555
ctg ttc tca ag gttggagcaa ttctcaactaa aattaatttggaaatgaaat		9542
Leu Phe Ser Arg		
560		
ctatcattta gtcacttggaa tgactttttt tatcagaaca taagcagggtt gtgtcttagtt		9602
ttcttttggtt gggtttagga cttaaaagtta tcttagtgta aaattttctc attttactaa		9662
accttaatgc ttatgttg ttgagttgc ag a ttg cac aac tag gga gag ctt		9716
Leu His Asn Gly Glu Leu		
565		
gta caa gag ata atc cac tgc taa gac caa gta tga gat ctt tag ttg		9764
Val Gln Glu Ile Ile His Cys Asp Gln Val Asp Leu Leu		
570	575	580
ttg ctc tta tga ccc ttt cat cac tta ctg agg att gtg atg atg aat		9812
Leu Leu Leu Pro Phe His His Leu Leu Arg Ile Val Met Met Asn		
585	590	595
ctt cct acg aaa gtc aaa ctc tca taa att tac tgt ctg tga gat aa		9859
Leu Pro Thr Lys Val Lys Leu Ser Ile Tyr Cys Leu Asp		
600	605	610
aggttctcca tgcaaatgca tgtttggat atatatcttg tagtacaact aagcagacaa		9919
aaagttttgtt actttgaatg taaatcgagt caggggtttt acattttattt actccatgt		9979
ttaattgccaa aaaccatcaa aaagtccatg gccagacttc ctgttaattt atttagcaaa		10039
gttgcagatt ctaagttcag ttttttata tataaggtttc agtattttt atatataatttta		10099
ttttataaaat ttttaactt gttacaatataa aacatattt gcattcatct tcaaatctt		10159

cagaatcaact tctcctacca cagaagctaa tagaagtgtc ttccagaatc aattcttcat 10219
ccactgtgaa aatctactat gtatcaaagc atgc 10253

<210> 24
<211> 621
<212> PRT
<213> Lotus japonicus Gifu

<400> 24

Met Lys Leu Lys Thr Gly Leu Leu Leu Phe Phe Ile Leu Leu Leu Gly
1 5 10 15

His Val Cys Phe His Val Glu Ser Asn Cys Leu Lys Gly Cys Asp Leu
20 25 30

Ala Leu Ala Ser Tyr Tyr Ile Leu Pro Gly Val Phe Ile Leu Gln Asn
35 40 45

Ile Thr Thr Phe Met Gln Ser Glu Ile Val Ser Ser Asn Asp Ala Ile
50 55 60

Thr Ser Tyr Asn Lys Asp Lys Ile Leu Asn Asp Ile Asn Ile Gln Ser
65 70 75 80

Phe Gln Arg Leu Asn Ile Pro Phe Pro Cys Asp Cys Ile Gly Gly Glu
85 90 95

Phe Leu Gly His Val Phe Glu Tyr Ser Ala Ser Lys Gly Asp Thr Tyr
100 105 110

Glu Thr Ile Ala Asn Leu Tyr Tyr Ala Asn Leu Thr Thr Val Asp Leu
115 120 125

Leu Lys Arg Phe Asn Ser Tyr Asp Pro Lys Asn Ile Pro Val Asn Ala
130 135 140

Lys Val Asn Val Thr Val Asn Cys Ser Cys Gly Asn Ser Gln Val Ser
145 150 155 160

Lys Asp Tyr Gly Leu Phe Ile Thr Tyr Pro Ile Arg Pro Gly Asp Thr
165 170 175

Leu Gln Asp Ile Ala Asn Gln Ser Ser Leu Asp Ala Gly Leu Ile Gln
180 185 190

Ser Phe Asn Pro Ser Val Asn Phe Ser Lys Asp Ser Gly Ile Ala Phe
195 200 205

Ile Pro Gly Arg Tyr Lys Asn Gly Val Tyr Val Pro Leu Tyr His Arg
210 215 220

Thr Ala Gly Leu Ala Ser Gly Ala Ala Val Gly Ile Ser Ile Ala Gly
225 230 235 240

Thr Phe Val Leu Leu Leu Ala Phe Cys Met Tyr Val Arg Tyr Gln
245 250 255

Lys Lys Glu Glu Lys Ala Lys Leu Pro Thr Asp Ile Ser Met Ala
260 265 270

Leu Ser Thr Gln Asp Ala Ser Ser Ser Ala Glu Tyr Glu Thr Ser Gly
275 280 285

Ser Ser Gly Pro Gly Thr Ala Ser Ala Thr Gly Leu Thr Ser Ile Met
290 295 300

Val Ala Lys Ser Met Glu Phe Ser Tyr Gln Glu Leu Ala Lys Ala Thr
305 310 315 320

Asn Asn Phe Ser Leu Asp Asn Lys Ile Gly Gln Gly Gly Phe Gly Ala
325 330 335

Val Tyr Tyr Ala Glu Leu Arg Gly Lys Lys Thr Ala Ile Lys Lys Met
340 345 350

Asp Val Gln Ala Ser Thr Glu Phe Leu Cys Glu Leu Lys Val Leu Thr
355 360 365

His Val His His Leu Asn Leu Val Arg Leu Ile Gly Tyr Cys Val Glu
370 375 380

Gly Ser Leu Phe Leu Val Tyr Glu His Ile Asp Asn Gly Asn Leu Gly
385 390 395 400

Gln Tyr Leu His Gly Ser Gly Lys Glu Pro Leu Pro Trp Ser Ser Arg
405 410 415

Val Gln Ile Ala Leu Asp Ala Ala Arg Gly Leu Glu Tyr Ile His Glu
420 425 430

His Thr Val Pro Val Tyr Ile His Arg Asp Val Lys Ser Ala Asn Ile
435 440 445

Leu Ile Asp Lys Asn Leu Arg Gly Lys Val Ala Asp Phe Gly Leu Thr
450 455 460

Lys Leu Ile Glu Val Gly Asn Ser Thr Leu Gln Thr Arg Leu Val Gly
465 470 475 480

Thr Phe Gly Tyr Met Pro Pro Glu Tyr Ala Gln Tyr Gly Asp Ile Ser
485 490 495

Pro Lys Ile Asp Val Tyr Ala Phe Gly Val Val Leu Phe Glu Leu Ile
500 505 510

Ser Ala Lys Asn Ala Val Leu Lys Thr Gly Glu Leu Val Ala Glu Ser
515 520 525

Lys Gly Leu Val Ala Leu Phe Glu Glu Ala Leu Asn Lys Ser Asp Pro
530 535 540

Cys Asp Ala Leu Arg Lys Leu Val Asp Pro Arg Leu Gly Glu Asn Tyr
545 550 555 560

Pro Ile Asp Ser Val Leu Lys Ile Ala Gln Leu Gly Arg Ala Cys Thr
565 570 575

Arg Asp Asn Pro Leu Leu Arg Pro Ser Met Arg Ser Leu Val Val Ala
580 585 590

Leu Met Thr Leu Ser Ser Leu Thr Glu Asp Cys Asp Asp Glu Ser Ser
595 600 605

Tyr Glu Ser Gln Thr Leu Ile Asn Leu Leu Ser Val Arg
610 615 620

<210> 25
<211> 623
<212> PRT
<213> Lotus japonicus Gifu

<400> 25

Met Lys Leu Lys Thr Gly Leu Leu Phe Phe Ile Leu Leu Leu Gly
1 5 10 15

His Val Cys Phe His Val Glu Ser Asn Cys Leu Lys Gly Cys Asp Leu
20 25 30

Ala Leu Ala Ser Tyr Tyr Ile Leu Pro Gly Val Phe Ile Leu Gln Asn
35 40 45

Ile Thr Thr Phe Met Gln Ser Glu Ile Val Ser Ser Asn Asp Ala Ile
50 55 60

Thr Ser Tyr Asn Lys Asp Lys Ile Leu Asn Asp Ile Asn Ile Gln Ser
65 70 75 80

Phe Gln Arg Leu Asn Ile Pro Phe Pro Cys Asp Cys Ile Gly Gly Glu
85 90 95

Phe Leu Gly His Val Phe Glu Tyr Ser Ala Ser Lys Gly Asp Thr Tyr
100 105 110

Glu Thr Ile Ala Asn Leu Tyr Tyr Ala Asn Leu Thr Thr Val Asp Leu
115 120 125

Leu Lys Arg Phe Asn Ser Tyr Asp Pro Lys Asn Ile Pro Val Asn Ala
130 135 140

Lys Val Asn Val Thr Val Asn Cys Ser Cys Gly Asn Ser Gln Val Ser
145 150 155 160

Lys Asp Tyr Gly Leu Phe Ile Thr Tyr Pro Ile Arg Pro Gly Asp Thr
165 170 175

Leu Gln Asp Ile Ala Asn Gln Ser Ser Leu Asp Ala Gly Leu Ile Gln
180 185 190

Ser Phe Asn Pro Ser Val Asn Phe Ser Lys Asp Ser Gly Ile Ala Phe
195 200 205

Ile Pro Gly Arg Tyr Lys Asn Gly Val Tyr Val Pro Leu Tyr His Arg
210 215 220

Thr Ala Gly Leu Ala Ser Gly Ala Ala Val Gly Ile Ser Ile Ala Gly
225 230 235 240

Thr Phe Val Leu Leu Leu Ala Phe Cys Met Tyr Val Arg Tyr Gln

245

250

255

Lys Lys Glu Glu Lys Ala Lys Leu Pro Thr Asp Ile Ser Met Ala
260 265 270

Leu Ser Thr Gln Asp Ala Gly Asn Ser Ser Ser Ala Glu Tyr Glu Thr
275 280 285

Ser Gly Ser Ser Gly Pro Gly Thr Ala Ser Ala Thr Gly Leu Thr Ser
290 295 300

Ile Met Val Ala Lys Ser Met Glu Phe Ser Tyr Gln Glu Leu Ala Lys
305 310 315 320

Ala Thr Asn Asn Phe Ser Leu Asp Asn Lys Ile Gly Gln Gly Gly Phe
325 330 335

Gly Ala Val Tyr Tyr Ala Glu Leu Arg Gly Lys Lys Thr Ala Ile Lys
340 345 350

Lys Met Asp Val Gln Ala Ser Thr Glu Phe Leu Cys Glu Leu Lys Val
355 360 365

Leu Thr His Val His His Leu Asn Leu Val Arg Leu Ile Gly Tyr Cys
370 375 380

Val Glu Gly Ser Leu Phe Leu Val Tyr Glu His Ile Asp Asn Gly Asn
385 390 395 400

Leu Gly Gln Tyr Leu His Gly Ser Gly Lys Glu Pro Leu Pro Trp Ser
405 410 415

Ser Arg Val Gln Ile Ala Leu Asp Ala Ala Arg Gly Leu Glu Tyr Ile
420 425 430

His Glu His Thr Val Pro Val Tyr Ile His Arg Asp Val Lys Ser Ala
435 440 445

Asn Ile Leu Ile Asp Lys Asn Leu Arg Gly Lys Val Ala Asp Phe Gly
450 455 460

Leu Thr Lys Leu Ile Glu Val Gly Asn Ser Thr Leu Gln Thr Arg Leu
465 470 475 480

Val Gly Thr Phe Gly Tyr Met Pro Pro Glu Tyr Ala Gln Tyr Gly Asp
485 490 495

Ile Ser Pro Lys Ile Asp Val Tyr Ala Phe Gly Val Val Leu Phe Glu
500 505 510

Leu Ile Ser Ala Lys Asn Ala Val Leu Lys Thr Gly Glu Leu Val Ala
515 520 525

Glu Ser Lys Gly Leu Val Ala Leu Phe Glu Glu Ala Leu Asn Lys Ser
530 535 540

Asp Pro Cys Asp Ala Leu Arg Lys Leu Val Asp Pro Arg Leu Gly Glu
545 550 555 560

Asn Tyr Pro Ile Asp Ser Val Leu Lys Ile Ala Gln Leu Gly Arg Ala
565 570 575

Cys Thr Arg Asp Asn Pro Leu Leu Arg Pro Ser Met Arg Ser Leu Val
580 585 590

Val Ala Leu Met Thr Leu Ser Ser Leu Thr Glu Asp Cys Asp Asp Glu
595 600 605

Ser Ser Tyr Glu Ser Gln Thr Leu Ile Asn Leu Leu Ser Val Arg
610 615 620

<210> 26
<211> 19
<212> DNA
<213> Lotus japonicus

<400> 26
aatgcttttg atcaggctg 19

<210> 27
<211> 20
<212> DNA
<213> Lotus japonicus

<400> 27
aggagcccaa gtgagtgcta 20

<210> 28
<211> 20
<212> DNA
<213> Lotus japonicus

<400> 28
caggaaaaac caccacctgt

20

<210> 29
<211> 21
<212> DNA
<213> Lotus japonicus

<400> 29
atggaggcga atacactgg g

21

<210> 30
<211> 1853
<212> DNA
<213> Lotus filicaulis

<400> 30
ttttctcttt ccctgttaac tatcatttgt tcccaacttc acaaacatgg ctgtcttctt 60
tcttacctct ggctctctga gtcttttct tgcactcactg ttgctttca ctaacatcgc 120
cgctcgatca gaacagatca gcggcccaga ctttcatgc cctgttgact caccccttc 180
tttgaaaca tatgtgacat acacagctca gtctccaaat cttctgagcc tgacaaacat 240
atctgatata tttgatatac gtccttgtc cattgcaaga gccagtaaca tagatgcagg 300
gaaggacaag ctggttccag gccaagtctt actggtacct gtaacttgcg gttgcgg 360
aaaccactct tctgccaata cctcctacca aatccagaaa ggtgatagct acgactttgt 420
tgcaaccact ttatatgaga accttacaaa ttgaaatac gtacaagctt caaaccagg 480
ggtaaatcca tatttgtgc cagagcgcgt caaagtcgtt ttccctttat tctgcagg 540
cccttcaaag aaccaggta acaaaggat tcagtatctg attacttatg tgtgaaagcc 600
caatgacaat gttcccttg tgagtccaa gtttggtgca tccccagcgg acatattgac 660
tgaaaaccgc tacggtaag acttcactgc tgcaaccaac cttccaaattt tgatcccagt 720
gacacagttt ccaaagctt ctaacaccc ttcaaatggg aggaaaagca gcattcatct 780
tctggttata cttggattt ccctggatg cacgttgcta actgcagtt taaccggac 840
cctcgatat gtatactgcc gcagaaagaa ggctctgaat aggactgctt catcagctga 900
gactgctgat aaactacttt ctggagttc aggctatgta agcaagccaa acgtgtatga 960
aatcgacgag ataatggaaag ctacgaagga tttcagcgat gagtgcagg ttggggatc 1020
agtgtacaag gccaacatag aaggtcggt tgtagcggt aagaaaatca aggaaggtgg 1080
tgccaaatgag gaactgaaaa ttctgcagaa ggtaaatcat ggaaatctgg tgaaactaat 1140
gggtgtctcc tcaggctatg atggaaactg tttctgggt tatgaatatg ctgaaaatgg 1200

gtctcttgcg gagtggctgt tctccaagtc ttcaggaacc ccaaactccc ttacatggtc 1260
tcaaaggata agcatagcag tggatgttgc tgtgggtctg caatacatgc atgaacatac 1320
ctatccaaga ataatacaca gggacatcac aacaagtaat atccttctcg actcgacctt 1380
caaggccaag atagcaaatt tcgccatggc cagaacttcg accaacccca tgatgcca 1440
aatcgatgtc ttgcgtttcg ggggtcttct gatagagttg ctcaccggaa ggaaagccat 1500
gacaaccaag gagaacggcg aggtggttat gctgtggaag gatatgtggg agatcttga 1560
catagaagag aatagagagg agaggatcag aaaatggatg gatcctaatt tagagagctt 1620
ttatcatata gataatgctc tcagcttggc atccttagca gtgaattgca cagctgataa 1680
gtctttgtct cgaccctcca tggctgaaat tggcttagc ctctccttgc tcactcaaca 1740
atcatctaac cccacattag agagatcctt gacttcttct gggtagatg tagaagatga 1800
tgctcatatt gtcacttcca ttacagcacg ttaagcaagg gaaggtaatt cag 1853

<210> 31
<211> 595
<212> PRT
<213> Lotus filicaulis

<400> 31

Met Ala Val Phe Phe Leu Thr Ser Gly Ser Leu Ser Leu Phe Leu Ala
1 5 10 15

Leu Thr Leu Leu Phe Thr Asn Ile Ala Ala Arg Ser Glu Gln Ile Ser
20 25 30

Gly Pro Asp Phe Ser Cys Pro Val Asp Ser Pro Pro Ser Cys Glu Thr
35 40 45

Tyr Val Thr Tyr Thr Ala Gln Ser Pro Asn Leu Leu Ser Leu Thr Asn
50 55 60

Ile Ser Asp Ile Phe Asp Ile Ser Pro Leu Ser Ile Ala Arg Ala Ser
65 70 75 80

Asn Ile Asp Ala Gly Lys Asp Lys Leu Val Pro Gly Gln Val Leu Leu
85 90 95

Val Pro Val Thr Cys Gly Cys Ala Gly Asn His Ser Ser Ala Asn Thr
100 105 110

Ser Tyr Gln Ile Gln Lys Gly Asp Ser Tyr Asp Phe Val Ala Thr Thr

115

120

125

Leu Tyr Glu Asn Leu Thr Asn Trp Asn Ile Val Gln Ala Ser Asn Pro
130 135 140

Gly Val Asn Pro Tyr Leu Leu Pro Glu Arg Val Lys Val Val Phe Pro
145 150 155 160

Leu Phe Cys Arg Cys Pro Ser Lys Asn Gln Leu Asn Lys Gly Ile Gln
165 170 175

Tyr Leu Ile Thr Tyr Val Trp Lys Pro Asn Asp Asn Val Ser Leu Val
180 185 190

Ser Ala Lys Phe Gly Ala Ser Pro Ala Asp Ile Leu Thr Glu Asn Arg
195 200 205

Tyr Gly Gln Asp Phe Thr Ala Ala Thr Asn Leu Pro Ile Leu Ile Pro
210 215 220

Val Thr Gln Leu Pro Lys Leu Thr Gln Pro Ser Ser Asn Gly Arg Lys
225 230 235 240

Ser Ser Ile His Leu Leu Val Ile Leu Gly Ile Thr Leu Gly Cys Thr
245 250 255

Leu Leu Thr Ala Val Leu Thr Gly Thr Leu Val Tyr Val Tyr Cys Arg
260 265 270

Arg Lys Lys Ala Leu Asn Arg Thr Ala Ser Ser Ala Glu Thr Ala Asp
275 280 285

Lys Leu Leu Ser Gly Val Ser Gly Tyr Val Ser Lys Pro Asn Val Tyr
290 295 300

Glu Ile Asp Glu Ile Met Glu Ala Thr Lys Asp Phe Ser Asp Glu Cys
305 310 315 320

Lys Val Gly Glu Ser Val Tyr Lys Ala Asn Ile Glu Gly Arg Val Val
325 330 335

Ala Val Lys Lys Ile Lys Glu Gly Gly Ala Asn Glu Glu Leu Lys Ile
340 345 350

Leu Gln Lys Val Asn His Gly Asn Leu Val Lys Leu Met Gly Val Ser
355 360 365

Ser Gly Tyr Asp Gly Asn Cys Phe Leu Val Tyr Glu Tyr Ala Glu Asn
370 375 380

Gly Ser Leu Ala Glu Trp Leu Phe Ser Lys Ser Ser Gly Thr Pro Asn
385 390 395 400

Ser Leu Thr Trp Ser Gln Arg Ile Ser Ile Ala Val Asp Val Ala Val
405 410 415

Gly Leu Gln Tyr Met His Glu His Thr Tyr Pro Arg Ile Ile His Arg
420 425 430

Asp Ile Thr Thr Ser Asn Ile Leu Leu Asp Ser Thr Phe Lys Ala Lys
435 440 445

Ile Ala Asn Phe Ala Met Ala Arg Thr Ser Thr Asn Pro Met Met Pro
450 455 460

Lys Ile Asp Val Phe Ala Phe Gly Val Leu Leu Ile Glu Leu Leu Thr
465 470 475 480

Gly Arg Lys Ala Met Thr Thr Lys Glu Asn Gly Glu Val Val Met Leu
485 490 495

Trp Lys Asp Met Trp Glu Ile Phe Asp Ile Glu Glu Asn Arg Glu Glu
500 505 510

Arg Ile Arg Lys Trp Met Asp Pro Asn Leu Glu Ser Phe Tyr His Ile
515 520 525

Asp Asn Ala Leu Ser Leu Ala Ser Leu Ala Val Asn Cys Thr Ala Asp
530 535 540

Lys Ser Leu Ser Arg Pro Ser Met Ala Glu Ile Val Leu Ser Leu Ser
545 550 555 560

Phe Leu Thr Gln Gln Ser Ser Asn Pro Thr Leu Glu Arg Ser Leu Thr
565 570 575

Ser Ser Gly Leu Asp Val Glu Asp Asp Ala His Ile Val Thr Ser Ile
580 585 590

Thr Ala Arg
595

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.